STUDIES ON THE EXPRESSION OF A CLONED *STAPHYLOCOCCUS AUREUS* PROTEIN A GENE IN *ESCHERICHIA COLI*

By

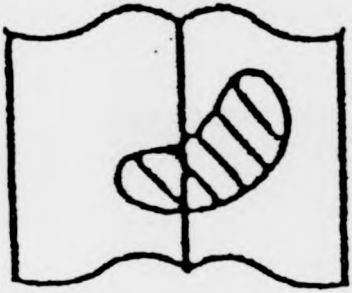
Helen L. Shuttleworth, BSc

This thesis is presented for the degree of Doctor of Philosophy in the Department of Biological Sciences, University of Warwick.

This work was carried out between 1983-1987 while employed at the Microbial Technology Laboratory, Centre for Applied Microbiology and Research, Porton Down, Salisbury, Wiltshire, SP4 0JG

JULY 1991

VARIABLE PRINT QUALITY



DEDICATION

TO MY MOTHER AND FATHER

CONTENTS

CHAPTER TWO MATERIALS AND METHODS

CHAPTER THREE NUCLEOTIDE SEQUENCE DETERMINATION OF THE SpA GENE

CHAPTER 4 EXPRESSION OF PROTEIN A IN *ESCHERICHIA COLI*

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS USED IN THIS THESIS

ADCC	Antibody dependent cellular cytoxicity
Ap	Ampicillin
bp	base pair
Bis	N,N-methylenebisacrylamide
BSA	Bovine serum albumin
C_H1	Constant region 1 on the immunoglobulin heavy chain
C_H2	Constant region 2 on the immunoglobulin heavy chain
C_H3	Constant region 3 on the immunoglobulin heavy chain
C_L	Constant region on the immunoglobulin light chain
CIC	Circulating immune complexes
CIP	Calf intestinal phosphatase
Cm	Chloramphenicol
dNTPs	Deoxyribonucleotides 5′-triphosphate
ddNTPs	Dideoxyribonucleotides 5′-triphosphate
DMF	Dimethyl formamide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DTT	DL-Dithiothreitol
DATD	N,N-Diallyltartramide
EDTA	Ethylene diamine tetraacetic acid
Elisa	Enzyme linked immonoabsorbant assay
Et Br	Ethidium Bromide
Fab	Monovalent antibody fragment of IgG
$F(ab)_2$	Divalent antibody fragment of IgG
Fc	Constant fragment of IgG
x g	Centripetal force equal to gravitational acceleration
g	Gram

G	Gibbs free energy
HgG	Human growth hormone
^{125}I	Radioactive Iodine
IGF-1	Human insulin-like growth factor
IgXg	Immunoglobulin of class X sub-group X
IPTG	Ispropylthio-β-galactosidase
Kb	Kilobase
Kc	Kilocalorie
kDa	Kilodaltons
KSCN	Potassium thiocyanate
l	Litre
M	Molar
MA	Milliamperes
mg	Milligrams
Min	Minute
ml	Millilitres
mM	Millimolar
mol	mole
mRNA	Messenger RNA
nm	Nanometers
OD_n	Optical density at wavelength n
^{32}P	Radiolabelled phosphate
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PEG	Polyethylene glycol
psi	Pounds per square inch
RF	Replicative form
RNase	Ribonuclease
rpm	Revolutions per minute

rRNA	Ribosomal ribonucleic acid
SDS	Sodium dodecyl sulphate
sec	Second
SpA	Staphylococcal Protein A
SpG	Streptococcal Protein G
STET	Sucrose, triton, EDTA, Tris buffer/lysis
T	Thymine
TBE	Tris borate EDTA buffer
TE	Tris EDTA buffer
TES	Tris EDTA sodium chloride buffer
TEMED	NNN'N'-tetramethylethylenediamine
tRNA	Transfer nucleic acid
Tween	Polyoxyethylenesorbitan monolaurate
u	Uracil
uv	ultraviolet light
v	volts
v_H	Variable region on the heavy immunoglobulin molecule
v_L	Variable region on the light immunoglobulin molecule
vol	Volume
(v/v)	Volume to volume ratio
(w/v)	Weight to volume ratio
X-gal (BCIG)	5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside
A	Angstroms

Other abbreviations used are as described in Biochem J. (1990). Vol. 265, p1-21.

ACKNOWLEDGEMENTS

I should like to thank Professor Tony Atkinson, my supervisors, Dr. Nigel Minton and Dr. George Salmond for their advice, encouragement and patience throughout the course of this work. In addition thanks are due to Dr. Clive Duggleby and Dr. Alan Warnes for the many interesting discussions that ensued throughout the SpA project.

I would also like to thank Roy Hartwell for oligonucleotide synthesis and N-terminal residue analysis of protein A. Dr. Alan Warnes, Dr. B. Dowsett for electron microscopy and Chris Barwis for aid with the statistical analysis of the SpA production results.

Thanks are also due to Sarah Bolam for typing this thesis. Also to Dr. Andrew Shuttleworth and Meg Sims for their unfailing support and to the many other colleagues who "proof-read" at very short notice.

DECLARATION

The work carried out in this thesis was the result of original research conducted by myself under the supervision of Dr. N.P. Minton and Dr. G.P.C. Salmond. Except for the Section on Detection of the Alkaline Phosphatase Gene which was carried out in collaboration with Dr. M.J. Taylor and was submitted in his PhD thesis entitled "Protein Engineering of Staphylococcal protein A".

SUMMARY

The gene coding for staphylococcal protein A has been cloned and the nucleotide sequence coding for the structural gene and its 5'flanking region determined. The DNA sequence showed the gene to be composed of a series of repetitive sequences, with five 120 nucleotide repeats comprising the 5'part of the structural gene and ten 24 nucleotide repeats in a continous sequence at the 3'terminus. Analysis of codon usage of the gene and comparison with other *S. aureus* genes revealed that *spa* showed a higher degree of homology with *S. aureus* plasmid genes than with chromosomally-located genes. Comparison with the reported codon preference in *E.coli* showed *spa* to have a similar bias, except for the *leu* codon.

An open reading frame of the *spa* structural gene showed 508 a.a residues, with a predicted protein molecular weight of 55,426. The N-terminal signal peptide comprised the first 36 amino acids. A fifth N-terminal domain (E) was shown to be homologous to the four IgG-binding domains previously reported to constitute the N-terminal portion of SpA.

The C-terminal part of SpA was shown to consist of two structurally different regions: an N-terminal repetitive region (Xr) composed of ten octapeptide repeats which are hydrophilic and have been proposed to span the Gram-positive cell wall, followed by a single domain (Xc), 58 a.a in length, whose C-terminal portion contains 20 hydrophobic residues, suggesting anchorage on the cell membrane.

Enhanced production of SpA in *E. coli* has been achieved by placing the *E. coli lac* promoter immediately upstream of the *spa* gene. This has allowed yields as high as 1.5g of SpA per litre of culture to be obtained from 150-litre batch fermenters. Problems were encountered during large-scale fermentation, as the recombinant clones were found to be extremely fragile when producing high levels of SpA.

Characterisation of the recombinant SpA indicated a molecular weight of 54,000 for the full length protein, compared with 56,000 for SpA produced in *S. aureus*. SDS-PAGE analysis of the IgG-binding proteins revealed extensive proteolysis of the recombinant SpA. Comparitive analysis of the IgG-binding polypeptides encoded by a truncated *spa* gene (lacking 69 a.a residues from the C-terminus), together with the determination of the N-terminal amino acid sequence, indicated that proteolysis occurred at the N-terminal portion of the Xc region. Cell fractionation studies and immunoelectronmicroscopy localised the SpA to the periplasm in *E. coli*. A possible regulatory control region for the *spa* gene has been localised upstream of the structural gene.

The *E.coli* alkaline phosphatase gene was isolated, cloned and its sequence determined to facilitate the production of SpA-alkaline phosphatase fusion proteins.

CHAPTER 1

1.1 *STAPHYLOCOCCUS AUREUS* PROTEIN A

1.1.1 Background

Staphylococcus aureus is a Gram-positive, facultatively anaerobic, non-motile bacterium of the Micrococcaceae family. The term "staphylococcus" reflects the characteristic grouping of the organism in irregular cell clusters (Greek staphyle, bunch of grapes), an individual coccus is 0.7 μm - 1.2 μm in diameter. Staphylococci are among the hardiest of the non-spore-forming bacteria. They can withstand dehydration for many weeks, many strains can survive temperatures as high as 60^{o}C for 30 mins, they are more resistant to disinfectants (such as mercuric chloride and phenol) than most bacteria and also have a very high salt tolerance. Of the three staphylococcal, species *S. aureus* is the most pathogenic and can cause a wide variety of suppurative diseases such as, superficial and deep abscesses, wound infections, septicaemia, endocarditis including metastatic abscesses, osteomylitis, pneumonia and emphysema, meningitis, and purulent arthritis; it also causes two forms of toxinoses: exfoliative skin disease and food poisoning (Davis *et al.*, 1980). Before the advent of antibiotics, staphylococcal bacteraemia resulted in an 80% mortality rate (Sheagren, 1984a). Since the 1950s, plasmid coded antibiotic resistance has increased, until now, 60%-90% of strains are resistant to two or more of the commonly used antibiotics and staphylococcal bacteraemia is once again one of the most common nosocomial infections (Sheagren 1984a, 1984b).

The Gram-positive cocci were first differentiated by Koch in 1878, and *S. aureus* as a species by Rosenbach in 1884 on the basis of colonial pigmentation (a golden yellow colouration, hence the name, "aureus"). In 1940 Verwey, using a serological test system to classify different strains developed an extraction procedure which resulted in five cellular fractions. Of these fractions, B was found to be highly antigenic, type specific and of a protein nature. In 1958 Jensen, classifying organisms using a gel precipitation method, found three antigens common to most strains of *S. aureus*. Of these, antigen A (classified as a polysaccharide) had strong antigenic properties and later experiments showed antibodies that reacted with antigen A to be present in normal

human serum, and in several samples of human colostrum and milk (Jensen, 1958a, 1958b, 1959). Lofkvist and Sjoquist (1962, 1963) and Grov *et al.*, (1964) showed that Jensen's antigen A was a protein, not a polysaccharide, and was in fact identical to the major component of Verwey's fraction B and the name protein A (SpA) was adopted. It was Forsgren and Sjoquist (1966) who first concluded that the reactivity of SpA to immunoglobulins was unusual in that it was directed against the Fc region of IgG as opposed to the normal binding of antigen to the antibody Fab sites.

1.1.2 Occurrence and production of protein A

Since the classification of SpA as a cell wall protein, several hundred strains of staphylococci and related organisms have been tested for the production of SpA (review by Langone, 1982a). SpA appears to be restricted almost entirely to *S. aureus*, exceptions being some strains of *S.hyicus* (Muller *et al.*, 1981) and *S.intermedius* (Lachia *et al.*, 1979) which also possess some immunoglobulin binding activity. Although no direct correlation between the production of SpA and the degree of pathogenicity of *S. aureus* has been shown to exist *in vivo* (Jonsson *et al.*, 1985), it has been shown to activate the complement system through the classical pathway. This may have implications for pathogenesis. It is also reported to cause histamine release, inhibit phagocytosis, inhibit antibody dependent cellular cytotoxicity (ADCC) and interact with the activation of B and T cells (reviews Langone, 1982b; Forsgren *et al.*, 1983). However, many of the reported results appear contradictory and may be caused by the dissimilar methodologies employed by different investigators. SpA has been found in 98.9% of *S. aureus* strains known to produce coagulase and DNase, the production of which is often correlated to pathogenicity (Forsgren, 1970, 1972). SpA has been shown to be produced in varying amounts by 95% of human *S. aureus* strains and 95% of animal *S. aureus* strains, the only exception being *S. aureus* isolated from bovine chronic mastitis where only 5% were SpA positive (Langone, 1982b). Sjoquist *et al.* (1972a) showed cell-bound SpA to be covalently linked to the peptidoglycan chains of the cell wall. Only 5% of the SpA positive strains produce predominantly

extracellular SpA, the majority of methicillin resistant strains have been shown to produce extracellular SpA (Lind, 1972; Winblad and Ericson, 1973; Movitz, 1976 and Lindmark *et al.*, 1977). Most of the strains producing cell-wall-bound SpA are also found to produce a small amount of extracellular SpA, up to 9% from exponentially growing *S. aureus* Cowan I, though higher yields are seen in the medium of stationary cultures. SpA has been shown to be produced during the exponential phase of cell growth and rapidly incorporated into the cell wall after synthesis, with production ceasing shortly after stationary phase is reached, followed by a significant fall in the amount of cell-bound SpA (Movitz, 1974). Therefore, high yields of extracellular SpA in predominantly cell-bound SpA strains during stationary phase have been attributed to autolysis rather than excretion. Although Forsgren (1969) found that as much as 10% of the total SpA was cytoplasmic, other investigators, (Movitz, 1974; 1976; Lindmark *et al.*, 1977) found only trace amounts in the cytoplasm (1%). Movitz (1976) showed that protoplasts from Cowan I produced the same total amount of SpA as intact bacteria, with 1% in the cytoplasm and 97% extracellular and therefore concluded that the bacterial cell wall was not necessary for *de novo* SpA synthesis. On the basis of SpA distribution, he also concluded that there may be no cytoplasmic stage involved in the production of SpA, i.e., SpA synthesised on membrane-bound ribosomes is directly incorporated into the cell wall.

1.1.3. Structure of protein A

To examine the structure of SpA it was first necessary to formulate reliable procedures for the isolation of the protein in a homogeneous and fully active form. The earliest attempts to isolate SpA involved boiling cells in various buffers, followed by precipitation with acid or high salt (Grov *et al.*, 1964; Grov, 1967; Lofkvist and Sjoquist, 1962; Forsgren and Sjoquist, 1969). Later attempts relied on enzymatic extraction using either lysozyme (Forsgren, 1969; Sjoquist *et al.*, 1972b) or lysostaphin (Sjoquist *et al.*, 1972a; Bjorck *et al.*, 1972). However, the majority of extraction procedures resulted in preparations exhibiting low SpA activity and which were composed of a heterogeneous collection of polypeptides, with molecular weights

ranging between 12 and 32 kDa and sedimentation coefficients between 1.26 and 2.6 (Langone, 1982b). Even with the homogeneous and fully active form of SpA produced using a lysostaphin extraction method, it was found that the molecular weight of the protein varied according to the determination method used. Thus, SDS-polyacrylamide gel electrophoresis gave values of 55 to 56 kDa, whereas chromatography and sedimentation techniques gave values of 41 to 42 kDa. The latter set of values are generally accepted as correct. This discrepancy in size determination may be caused by the unusual configuration of SpA. This was first alluded to by Bjorck (1972) whose hydrodynamic studies of SpA, released from Cowan I using lysostaphin, revealed a fractional ratio of 2.1 and 2.2 and an intrinsic viscosity of 29 ml/g. These parameters were interpreted as indicating an extended shape rather than a typical globular protein. Circular dichroism studies by Sjoholm (1975) were consistent with the proposed extended shape and indicated an ordered structure composed of 50% right-handed helical regions and 10% to 20% β-structure.

Amino acid analysis of intact Cowan I SpA isolated by lysostaphin treatment, indicated that it comprised some 378 amino acids, but contained no tryptophan or cysteine residues (Sjoquist *et al.*, 1972b). Further analysis of lysostaphin isolated Cowan I SpA by Movitz (1976) and Lindmark *et al.* (1977) also showed no tryptophan or cysteine residues to be present, however, the total number of amino acids were slightly different, 381 and 395 respectively.

Sjoquist *et al.* (1972b) also analysed the precipitation reaction between SpA and human IgG. They concluded that one molecule of SpA was able to bind two molecules of IgG, and suggested that SpA might consist, in part, of repeating units. The existence of the SpA1-IgG2 complex was confirmed by Mota *et al.* (1978), who went on to suggest, on the basis of further precipitation reactions of SpA1-rabbit IgG2 complex with rabbit IgG, sheep erythrocytes and human IgG, that the SpA molecule may contain four combining sites with different affinities for IgG. Hjelm *et al.* (1975) confirmed the

repetitive structure of SpA, following the amino acid analysis and mapping of purified SpA fragments generated by partial tryptic digestion, capable of binding to the Fc region of human IgG. Their results suggested the presence of at least three similar regions in SpA. A similar series of studies were undertaken by Sjodahl (1976; 1977a; 1977b). SpA fragments, derived by both lysostaphin and trypsin digestion, were purified by affinity chromatography on IgG Sepharose and their amino acid sequence determined using manual Edman degradation. On the basis of this data, a model for the structure of SpA was proposed, suggesting that the molecule is composed of two structurally and functionally different regions. (i) an N-terminal immunoglobulin-binding region, with a molecular weight of 27 kDa, composed of four consecutive units, D, A, B, and C, and; (ii) a C-terminal region which exhibited no immunoglobulin binding abilities.

A partial amino acid sequence of the C-terminal region, designated X, revealed that it consisted of an octapeptide sequence repeated several times. The inability of trypsin to generate region C material when used on intact, rather than on lysostaphin treated cells may be explained by the region X being buried in the cell wall, making the trypsin site separating region C from region X inaccessible. Amino acid sequence of the N-terminal immunoglobulin-binding units showed them to be between 58 to 60 amino acids in length with an 80% homology level between the repeat units, as well as internal homology within each unit. Sjoquist *et al.* (1972a) had showed that region D was preceded by another region, but a blocked N-terminus made amino acid analysis difficult. The subsequent determination of the nucleotide sequence of the N-terminal region of the *spa* gene of strain 8325-4 (Lofdahl *et al.*, 1983) confirmed the presence of a coding region 5′ to that encoding region D. The predicted amino acid sequence of the SpA N-terminus was characterised by the presence of an atypically long signal peptide, followed by 50 amino acids highly homologous to those of the four previously characterised immunoglobulin regions. This new domain was designated region E. Lofdahl *et al.* (1983) also noted the presence of a possible promoter region and ribosomal binding sequence upstream of the initiation codon.

The structure of SpA, in particular the nucleotide sequence, has been further elaborated upon within this thesis (strain Cowan I, NCTC8530) and also in a series of papers (strain 8325-4) published during the course of the practical experimentation work and writing of this thesis. These papers are discussed in Chapter 3, where comparison of the nucleotide sequences is undertaken.

1.1.4 The interaction between protein A and immunoglobulins

The early studies of Forsgren and Sjoquist (1966) had shown that the immune interaction between IgG and protein A was non-specific, with the protein A binding to the Fc-region of the IgG molecule and not the Fab-region. A more precise localisation of the protein A reactivity was subsequently obtained by Kronvall and Frommel (1970), when examination of isolated $F(ab')_2$, Fab, Fc and Fc' fragments, along with pepsin component II and III (see Figure 1.1), showed the binding site for SpA to be located on pepsin component II; this corresponds mainly to the C_H2 domain of IgG.

Attempts to further identify the exact region of the Fc fragment involved in the binding reaction between IgG and SpA, and the amino acid residues involved have been carried out using chemical modification techniques, but with only a limited amount of success. Results did, however, suggest that the presence of covalently linked C_H2 and C_H3 were necessary for protein A binding (Endresen and Grov, 1976; Stewart *et al.*, 1978; Tarkhanova *et al.*, 1980). X-ray crystallographic investigations carried out by Deisenhofer *et al.* (1978) and Deisenhofer (1981), have defined both the secondary structure of a purified fragment B of protein A and the Fc fragment of IgG and the nature of the interaction between these two fragments. The C_H2 and C_H3 domains of the Fc fragment of IgG consists of 2 layers of anti-parallel β-pleated sheets, enclosing a mainly non-polar interior. The C_H2 chains are more disordered than the C_H3 chains as they do not form a lateral contact. This dissimilarity between C_H2 and C_H3 gives a flexible interface between the two regions.

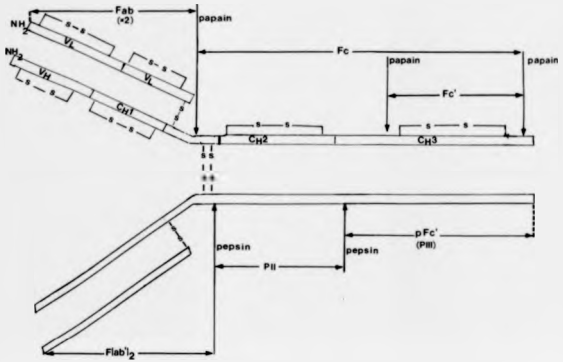


Figure 1.1 Diagrammatic representation of the structure of human IgG1

This figure was adapted from Edelmann *et al.*, (1969) and shows the different domains of the heavy (H) and light (L) chains; the variable regions (V) and the constant regions (C). The intrachain and interchain disulphide bonds are shown. The cleavage sites for the proteolytic enzymes papain and pepsin are indicated, the various peptide fragments produced and their nomenclature are marked.

Fragment B of protein A is a small globular protein consisting of three helices which form a regular triangular array. There are two antiparallel α helices consisting of three turns each, and a C-terminal helix consisting of two turns, with the remainder of the molecule being elongated.

The data of Deisenhofer (1981) indicated two contacts between fragment B and Fc, however, it is thought that only one is normally formed under physiological conditions and the other is a crystal contact only. The former contact, which is predominantly hydrophobic, involves residues from both the N-terminal helices and the C_H2 and C_H3 domains of Fc. Investigations by Mota *et al.* (1978) and Langone *et al.* (1978 a&b), have shown that a complete protein A molecule, containing IgG binding regions E, A, B, C and D, can bind two IgG molecules. More recent studies (Hanson and Schumaker, 1984; Moks *et al.*, 1986) indicate that IgG is functionally bivalent for SpA, which itself is at least tetravalent and possibly even pentavalent. Hanson and Schumaker propose that the composition of the IgG-SpA complex varies according to the ratio of IgG and SpA molecules present, with an IgG_1-SpA_1, and IgG_2-SpA_2 complex being favoured at high SpA ratios (4:1 and above) and a IgG_4-SpA_2 complex being favoured at higher IgG to SpA ratio (1.7:1) to (4:1).

In addition to an Fc binding ability, SpA also exhibits a weak reactivity to the Fab fragment of many classes of immunoglobulins (Endresen, 1979; Inganas *et al.*, 1980). This interaction does not involve the antigen combining site and is also able to act independently of the Fc activity of a single IgG-SpA domain (Inganas *et al.*, 1981). It has been shown that in order to precipitate SpA, and for complement activation by SpA, the IgG involved must have both Fc and Fab reactivities, (Endresen, 1979; Inganas and Nilsson, 1981).

The binding specificities of protein A for the Fc fragment of the different types of immunoglobulin and their sub-classes are summarised in Table 1.1. With few exceptions SpA reacts strongly with the IgG of almost all mammals. Notable

TABLE 1.

Binding Specifities of Protein A to the Fc Fragment of the Immunoglobulin

Species	Type of Immunoglobulin	Subgroup
Man	IgG	1,2,4, (3 variable)
	IgA	2
	IgM	2
Rabbit	IgG (Soluble complex)	-
Mouse	IgM (weakly)	-
	IgG	1 (weakly) 2a, 2b, 3
Rat	IgG	1, 2c
Guinea pig	IgG	1, 2
Cow	IgG	2 (weakly)
Sheep	IgG	2 (weakly)
Goat	IgG	2 (weakly)
Horse	IgG	a, b, c (T)
Dog	IgG	a, b, c, d
	IgA (some)	
	IgM (some)	

Protein A does not bind Avian IgG, though it has been reported to bind to small amounts of chicken IgM (3%).

The table was adapted from review articles by Langone (1982a) and Surolia *et al.*, (1982).

exceptions are ruminant IgG and subclass 3 of human IgG (Langone, 1982a). The failure of most human IgG 3 immunoglobulins to bind to SpA is attributed to the presence of an amino acid change at position 435 in the Fc fragment, where arginine has been substituted for histidine. When the substitution was included in the models of Deisenhofer (1981) it was shown that arginine would considerably affect the normal Fc-SpA complex as the arginine side chain could form additional contacts with other amino acids of the complex, while its positive charge would also remain unbalanced. Such major affects would prevent arginine fitting into the normal Fc-SpA complex. Other studies have shown that SpA can bind to certain human IgA's and IgM's, Table 1.1 (reviewed by Langone, 1982a). SpA also binds to some IgA and IgM groups in some mammalian species.

Initial reports on the binding of SpA to the immunoglobulins of non-mammalian species (Kronvall *et al.*, 1970; 1974) indicated that only the serum from two primative flightless birds (*Rhea americana* and *Pterocnemia pennata*) showed positive reactions. No reactions were recorded among Amphibian, Reptilian, Fish or among all other birds tested. However, Zikan *et al.* (1980) reported that small amounts of non-mammalian serum IgM of chicken, clawed toad and carp bound to SpA. Subsequent tests on purified carp IgM appeared to locate the majority of the binding activity to the Fab region. Although protein A has been reported to bind weakly to the Fab fragments of some IgG and IgE globulins. These particular IgM and Fab-SpA binding results were considered dubious by Langone (1982a) who concluded that the binding seen may have been due to Fab antibody activity against SpA.

1.1.5 Uses of protein A

The ability of protein A to bind to immunoglobulins without inhibiting the antigen-antibody reaction has resulted in its use in a wide range of analytical and preparative immunological techniques, and a more limited number of therapeutic applications (reviews: Goding, 1978; Langone, 1982a, 1982b).

1.1.5.1 Analytical and preparative immunological techniques

One of the most significant advances in immunological techniques was the formulation by Hjelm *et al.* (1972) of a reliable, low cost method for purifying immunoglobulins from serum and ascites fluid, using affinity chromatography on protein A Sepharose. In this simple one step procedure, immunoglobulin can be recovered from the column using a variety of techniques including; lowering or raising the pH (Hjelm *et al.*, 1972); the use of chaotropic buffers such as KSCN, (Kronvall, 1973), or electrophoretic elution (Morgan *et al.*, 1978). The use of SpA-Sepharose columns provides not only rapid high performance affinity purification of IgG immunoglobulins *per se*, but also facilitates the isolation of immunoglobulins for the subsequent purification of antigens. Furthermore, they may be employed in the analysis of crude biological samples for the presence of monoclonal antibodies.

SpA has been labelled with many types of tracer, including radionucleotides, enzymes fluorescent probes and metals (Langone, 1982b). Labelled SpA is often used in direct binding assays where it replaces the second antibody. Examples of some of the uses of labelled SpA are listed below:

(A) The study of surface antigens on intact cells is greatly enhanced by the use of labelled SpA as an anti-δ reagent in place of the normal second antibody (anti-receptor antibody) (Goding, 1978). Its use avoids the problems of lack of specificity of antisera and the non-specific binding of the anti-receptor antibodies to the cells via Fc receptors, observed by Warner (1974).

(B) The quantitative determination of antibodies and antigens either on the cell surface or in solution, for a wide variety of purposes:

(i) Monitoring the production of monoclonal antibodies by lymphocyte hybridomas. In particular SpA labelled with ^{125}I followed by autoradiography allows a rapid and inexpensive screening procedure for several hundred hybridoma cells

(Brown *et al.*, 1979).

(ii) Evaluation of polyclonal B cell activation by various stimulatory substances using the haemolytic plaque assay developed by Gronowicz *et al.* (1976).

(iii) The detection and quantification of antigens and IgG antibodies on human lymphoid cells, following rosette formation between these cells and SpA (labelled with glutaraldehyde) coated erythrocytes (Ghetie *et al.*, 1984).

(iv) Labelled SpA either in the form of $^{[32]}P$ labelled *S. aureus* or $^{[125]}I$ SpA can be used in an assay for the quantitation of circulating immune complexes (CIC). The CIC are precipitated from serum using polyethyleneglycol and the washed precipitate either incubated with labelled SpA (Stevens and Bridis, 1981) or the CIC precipitated, redissolved and absorbed to *S. aureus* (via SpA receptors) and stimulated by incubation with $^{[125]}I$ SpA, (Barkas, 1981). The latter method relies on the availability of multiple SpA binding sites. This type of approach will not detect all complexes, but is superior to the alternative platelet aggregation test as it can be carried out accurately on older serum samples whereas stored platelets do not aggregate.

(C) SpA labelled with a heavy metal tracer (i.e. gold) can be used in conjunction with electron microscopy to visualise the localisation of a protein to a particular cellular compartment within sectioned cells (Bendayan, 1984; Baron, *et al.*, 1989).

The use of SpA in place of a second antibody offers several advantages; it requires only one tracer regardless of the antigen used, non-specific binding is lower, it is as sensitive as the classical assay method while having a quicker reaction time, it can be used in situations where the conditions necessary to introduce other forms of tracer may prove harmful to labile substances, and unlike many second antibodies whose activity is often severely affected by the different methods of labelling, SpA retains a high degree of functional activity when labelled with a variety of reagents.

The utility of SpA as an immunological tool has been considerably enhanced by the application of recombinant techniques. Using such technology gene fusions may be made composed of DNA encoding SpA and sequences specifying the enzyme conjugate to act as the tracer. *In vivo* translation of the constructed hybrid "gene" by a bacterial cell results in a multifunctional chimaeric protein, containing the SpA IgG binding regions and the desired enzyme activity. Fusion proteins produced in this manner have several advantages over SpA-enzyme conjugates made by chemical means. Most notably a gene fusion system will yield a defined conjugate in terms of attachment site and molar ratio between carrier and peptide. The loss of enzyme activity often associated with the process of chemical conjugation will be avoided. Furthermore, when a synthetic peptide is being used as a tracer, it is easier to manufacture using oligodeoxynucleotide synthesis (which allows for rapid and efficient production of up to 150 bp, 50 a.a) than by peptide synthesis. The latter route relies on more complex methodology, which is both time consuming, and often fails as some peptide sequences are inherently difficult to synthesise.

The subsequent purification of the recombinant SpA fusion proteins is readily achieved in a single affinity chromatographic step by capitalising on the IgG binding properties of the SpA moiety (Uhlen *et al.*, 1984a). For some immunological procedures, such as acting as an antigen for antibody generation the eluted fusion protein is suitable for immediate use. Further processing may be required for direct structural or functional studies. In the former case, the SpA region does not interfere with the immune response. Indeed, Lowenadler *et al.* (1986; 1987), describes a novel method for obtaining specific antibodies against short peptides in which the repetitive structure of SpA acted as an adjuvant enhancing the immune response against the attached short peptide. The system described by Lowenadler *et al.* (1986; 1987) has also been used to express longer synthetic genes from which antibodies were successfully prepared. This should allow for the development of antibodies directed against both linear and discontinuous epitopes, the latter accounting for a substantial fraction of the antibodies produced against a variety of native proteins (Berzofsky, 1985). It should also be noted

that the SpA- fusion protein used to induce the production of polyclonal and monoclonal antibodies can also be used for the screening of the resulting hybridoma (Valerie *et al.*, 1987).

The use of the purified fusion protein in structural and functional studies often requires the separation of the SpA affinity tail from the protein of interest. This separation can be achieved by introducing a specific chemical or enzymatic cleavage site at the junction between the two proteins. Nilsson *et al.* (1985b), used such a system to facilitate the production and purification of native human insulin-like growth factor (IGF-1). A synthetic IGF-1 gene was fused to SpA and then, using site directed mutagenesis, an acid labile Asp-Pro cleavage site introduced at the fusion point. Following recombinant production, the fusion protein was eluted from an affinity column and treated with 70% formic acid to cleave the Asp-Pro peptide bond. The mixture of the two cleaved proteins was then passed through another affinity column, where the SpA portion of the fusion and the non-cleaved materials bound to the IgG ligand and were retained, leaving the desired protein to be collected in the eluate.

To further enhance the use of *spa* in gene fusions a number of protein fusion vectors have been developed by Nilsson *et al.* (1985a, 1985b). These vectors contain DNA encoding the IgG binding region of SpA followed by a multiple cloning site containing 5 unique restriction sites for insertion of a foreign gene, allowing for rapid construction of the required gene fusion.

1.1.5.2 Therapeutic applications

The anti-tumour effects of some Gram-positive microorganism extracts were reported as early as 1906 (Coley, 1906). In more recent years it was observed that after surgery to remove lung cancer, patients who had a staphylococcal emphysema had a better survival rate than those who had an uncomplicated post-surgical period (Ruckdeschel *et al.*, 1972, Takita, 1970). A potential role for SpA became more apparent when it was

demonstrated that circulating immunoglobulins appeared to act as blocking factors, inhibiting cell-mediated cytotoxicity against tumour cells (Baldwin *et al.*, 1973; Hellstrom *et al.*, 1969; Hellstrom and Hellstrom, 1970; Sjorgren *et al.*, 1971). Circulating immunoglobulin was also found in abnormally high levels in patients with acute immune diseases (Jones *et al.*, 1986; Steel *et al.*, 1974). Therefore, SpA appeared a logical host reagent for removing immune complexes from tumour-bearing serum by affinity binding extracorporeally, allowing the later return of the plasma to the patient. The only other method of removing the circulating immune complexes would be the removal of the plasma, followed by return of donor plasma, with the associated risk of contaminated blood products, or the use of a plasma replacement solution of albumin and saline, which lacks essential plasma components. The fact that SpA has been shown to have a higher affinity for immunocomplexes than for free IgG allows for removal of the damaging immunocomplexes while allowing the plasma to retain a reasonable level of free IgG (Kessler, 1975). However, it soon became apparent that the immunological consequences of plasma therapy were far more complex, with many researchers noting the occurrence of tumour regression following the passage of plasma volumes too small to achieve significant blocking factor removal (reviews: Fer and Oldham, 1985; Mackintosh *et al.*, 1985; Jones *et al.*, 1986 and Solal-Celigny, 1985). Romagnani *et al.* (1980), observed that serum CIC levels showed a similar decrease in dogs with mammary carcinoma (in both those exhibiting tumour regression and those not). Therefore, the removal of a significant portion of circulating CICS did not appear to be essential for tumouricidal effects; although there was no doubt that perfusion of patient serum over protein A bearing *S. aureus*, and over purified SpA, resulted in clinically significant responses. The possibility of direct activation of hormonal antitumour effects through SpA interactions with sera from tumour bearing animals was suggested by Steele *et al.* (1974). Further experiments then ensued, which although supporting the activation of hormonal antitumour effects, cast doubt on the role of SpA in the interaction. These often contradict one another. For instance in *in vitro* experiments with SpA treated plasma used against acute myeloid leukaemia blast cells (AML), it has been shown that (i) the plasma must come from AML patients to obtain

cytotoxicity and (ii) the treated sera is non-toxic to normal cells and that SAW (*S. aureus*, Wood 46) not bearing SpA did not induce toxicity (Miller *et al.*, 1982). However, in *in vivo* studies on dogs undertaken by Gordon *et al.* (1983) on canine spontaneous tumour systems, it was shown that perfusion of plasma over SpA producing *S. aureus* cells, or over non-SpA bearing SAW, was equally effective in causing regression of tumours in 6 out of 9 dogs. If SpA is the active agent, direct infusion might elicit similar tumouricidal effects. While such cases have been reported (Ray and Bandyopadhyay, 1983) using relatively pure SpA, a series of animal studies by Cohen *et al.*, (1984) showed no tumouricidal effects and it is generally accepted that the interaction of SpA with plasma *in vivo* results in no antitumour effects.

The role of SpA has been further questioned by the work of Sukumar *et al.* (1984). In these experiments, rats treated with plasma which had been exposed to inactivated CNBr-Sepharose showed comparable reductions in tumour size to those treated with plasma exposed to SpA covalently linked to Sepharose. This suggested that the presence of SpA may not be the critical factor. Further clinical studies (reviewed by Fer and Oldham, 1985) to duplicate the very effective anti-cancer treatment using SpA described by Terman *et al.* (1981), have failed to elicit such dramatic results, even when an identical materials and methods procedure was followed (Fer *et al.*, 1984). The only difference that could be traced was that the method used to produce the SpA by Pharmacia (supplied in both studies) had been modified to increase the yield and purity. This led to the suggestion (Fer *et al.*, 1984) that other bioactive staphylococcal products may have previously co-purified with protein A, and that these had been eliminated in the later Pharmacia product.

The role of SpA in perfusion systems is, therefore, controversial with some experimenters indicating that perhaps another component from *S. aureus* is the active factor, while others go even further and suggest that no bacterial product is necessary for the tumouricidal effects seen. The only common denominator necessary for tumour

regression in all the studies being the need for plasmaphoresis of tumour-bearer plasma. Fer and Oldham (1985) have suggested that perhaps a broad range of materials may trigger certain mechanisms mediated by a plasma factor, and that it is only through a series of detailed and systematic studies that the role of plasma immunoadsorption, and thus the role of SpA in cancer therapy, if any, will become apparent.

Although the direct treatment of cancer using SpA may not be effective, SpA plasmaphoresis can still be a vital part of cancer therapy. SpA can be used to counter some of the side effects associated with some standard chemotherapy treatments; such a use is reported by Zimmerman *et al.*, (1982) for the treatment of a tumour associated syndrome similar to thrombotic thrombocytopenic purpura. This syndrome arose following a rapid response in a patient treated with the chemotherapeutic drugs, fluorouracil, doxorubicin and mitomycin. It was hypothesised that as a tumour is reduced by chemotherapy the pre-existing state of antigen excess decreases, allowing the formation of soluble antigen-antibody complexes. These complexes induce the release of platelet substances which have the potential to initiate intramuscular coagulation and, because they accumulate in the blood vessels, cause local vascular injury, as well as initiating the deposition of platelets and fibrin, leading to the above syndrome. Plasmaphoresis was used in 5 treatment stages to reduce the CIC level of 324 mg l^{-1} to within the normal range of 10 to 65 mg l^{-1}. This particular use for SpA may become more common as more effective chemotherapeutic drugs with greater potency are developed and a rapid response by the patient becomes a more common occurrence.

Plasmaphoresis using SpA-Sepharose can also be used as a procedure for removing high titre antibodies in short term therapeutic procedures. An example of such a case, where the role of SpA-Sepharose was vital, has been reported by Nilsson *et al.* (1981). A haemophiliac with a high titre of factor IX antibody required surgery for an invasive pseudotumour into the base of his left elbow. Exchange plasmaphoresis was not feasible because of the possibility of stimulating factor IX inhibitor, and the fact that

the patient had antibodies against tissue antigens. It was not possible to neutralise the anti-factor IX antibodies using a factor IX concentrate, as the commercially available products contain small amounts of activated coagulation products, which in the amount necessary to treat the above case could have lead to intravascular coagulation. Treatment of 6000 ml of the patients plasma to remove antibodies by extracorporeal adsorption of the plasma to SpA-Sepharose, followed by transfusion, decreased the antibody titer and total immunoglobulin content to 20% of the original values. This allowed the remaining antibodies to be neutralised by infusion of factor IX concentrate. Conventional substitution therapy, combined with immunosuppression to prevent further antibody synthesis, was then given. This procedure allowed the operation to be performed without further complications.

1.2 Choice of Host System for High-Level Expression of SpA

Recent developments in biological science dependent on the IgG binding properties of purified SpA or SpA-fusions have, of necessity, increased the requirement for the amount of purified SpA and SpA-fusion products. The normal method of protein A production and purification from a SpA producing strain of *S. aureus* becomes a limiting factor both in terms of safety and the complexity of the recovery process. As *S. aureus* is pathogenic and produces several endotoxins, large batch cultures must be grown and harvested under strict safety conditions. These rules also apply to *S. aureus* Cowan I, considered one of the least pathogenic strains of *S. aureus*. Recovery of SpA from *S. aureus* requires the use of lysostaphin to release the SpA from the Gram-positive cell wall, where it is covalently linked to the peptidoglyan structure. In *S. aureus* strain Cowan I expression of the chromosomally located *spa* gene results in the production of SpA at a level of 1.7% dry weight of the bacteria (Sjoquist *et al.*, 1972b).

Primarily, *S. aureus* is not suitable for large scale production of protein A on grounds of pathogenicity. An alternative host combined with a suitable plasmid vector system

offers greater potential for increasing the production of SpA. The relative merits of using either a Gram-negative or Gram-positive bacterium as a host will be considered in the next two sections.

1.2.1 Gram Positive Host

The two Gram-positive bacteria considered as possible hosts for the expression of the SpA gene were; (i) a non-pathogenic staphylococcal species such as *Staphylococcus xylosus*, which is normally located on the human skin (Schleifer and Kloos, 1975; Uhlen *et al.*, 1984b), or, (ii) *Bacillus subtilis*, a non-pathogenic, aerobic, soil bacterium, which is often grown in large scale fermentations in order to isolate one of the 30 natural exo-proteins and is well documented (reviews Priest, 1977; Henner and Hoch, 1980 and Young, 1980).

1.2.1.1 The potential of a staphylococcal host

The main advantage in using a staphylococcal host such as *S. xylosus* is the high chance of recognition of the *spa* gene transcription and translation regulation sequences and correct processing of the SpA polypeptide to yield an authentic native protein. There are several disadvantages in using *S. xylosus*. Of the few staphylococcal plasmids that are well characterised, none were isolated from *S. xylosus*. Any recombinant plasmid studies undertaken on staphylococcal species requires the use of a protoplast transformation system (Gotz *et al.*, 1981); inherently more difficult than the transformation system used in Gram-negative *E. coli*. There is also the possibility that *S. xylosus* may contain a gene regulation mechanism that will limit the production of SpA. Uhlen *et al.* (1984b) demonstrated that *S. xylosus* did not produce a protein A type protein. It is thought unlikely to carry a gene encoding SpA, as although SpA proteins have been reported for other staphylococcal species, these have been coagulase positive species such as *Staphylococcus hyicus* (Muller *et al.*, 1981) and *Staphylococcus intermedius* (Lachia *et al.*, 1979). No coagulase negative *Staphylococci*

have been reported to produce IgG binding proteins, though they have been shown to produce other serum precipitating proteins (Langone, 1982a and Osland, 1981). However, *S. aureus* Wood 46, a strain thought not to produce SpA, has been shown to contain an *spa* gene within its chromosome whose transcription appears to be completely repressed (C. Duggleby, personal communication); Uhlen *et al.*, 1984b). Uhlen *et al.* (1984b) have also shown that the expression of the plasmid encoded protein A gene in various strains of *S. aureus* is strongly correlated to the expression of the chromosomal *spa* gene, and that plasmid encoded *spa* when expressed in a staphylococcal coagulase-negative species is severely limited, producing only 12 to 30% of the SpA coded for by the same gene when chromosomally located in *S. aureus* strains Cowan I and A676.

Therefore, it would seem unlikely that any non-pathogenic species of staphyloccoci would prove a suitable host for an expression vector capable of producing large amounts of SpA.

1.2.1.2 The use of *Bacillus subtilis* as the recombinant host

B. subtilis has a wide variety of well characterised and well used cloning vectors, the majority of which have been derived from *S. aureus* plasmids (Ehrlich *et al.*, 1982, Gryczan, 1982). *B. subtilis* also has the ability to export many proteins to the exterior of the cell (Priest, 1977). *B. subtilis* shows a low bias in codon usage similar to that of *S. aureus*. Being Gram-positive it will have a similar cell wall structure to *S. aureus*, this may facilitate the correct localisation of a normally exported *S. aureus* gene product. Several *S. aureus* genes have been cloned into *B. subtilis* and expressed with varying degrees of success, indicating that *B. subtilis*, is able to correctly recognise some *S. aureus* secretion signals i.e. *S. aureus* α-haemolysin is correctly exported to the cell supernatant (Fairweather *et al.*, 1983). The *S. aureus* gene encoding β-lactamase has been successfully expressed when integrated into the *B. subtilis* chromosome (1% total cell protein). However, the β-lactamase produced was found to be cell-associated

rather than secreted to the cell exterior as expected (Saunders *et al.*, 1984).

A major problem with using *B. subtilis* for genetic manipulation is that many of the plasmids, in particular high copy number plasmids and recombinant plasmids, are extremely unstable when in *B. subtilis*. Two types of instability have been recognised firstly, segregational instability, which means, the loss of the entire plasmid from the host cell (Kreft *et al.*, 1982; Yoshimura *et al.*, 1983). Secondly, structural instability, which involves rearrangement of plasmid sequences frequently involving the formation of deletions (Gryczan and Dubnau, 1978; Kreft *et al.*, 1982). Some genes that could not be established on multicopy number plasmids in *B. subtilis* have been effectively expressed when integrated into the *B. subtilis* chromosome (Saunders *et al.*, 1984). However, such strains will only carry a single copy of the gene and the levels of heterologous protein produced may for this reason, be severely limited.

Subsequent parallel studies in this laboratory which attempted to introduce the *spa* gene into *B. subtilis* using multicopy number plasmids based on the *B. subtilis* replicons pUB110 and pC194 (Ehrlich *et al.*, 1982) failed to yield a stable transformant that expressed SpA (L.E. Clarke, personal communication). Further investigations revealed that major deletions and rearrangements of the *spa* sequence encompassing the promoter region and domains E, A and D of the IgG binding regions had occurred. Subsequent results published by Fahnestock *et al.* (1986a) showed that stable transformants carrying the entire *spa* gene could be obtained by integration of *spa* into the chromosome. However, using a multicopy plasmid system based on either the pUB110 or the pC194 replicon Fahnestock *et al.* (1986a) reported a similar lack of success in cloning the entire *spa* gene, but showed that the coding sequence of the native protein could be stably maintained as a pUB110 based plasmid providing the *spa* expression signals were replaced by an expression signal less active during growth. In this particular case Fahnestock *et al.* (1986a) replaced the *spa* promoter and translational start codon with the promoter region from a *Bacillus amyloliquefaciens* α-amylase gene. A similar plasmid stability problem was encountered by Saunders *et al.*,

(1984), when trying to express the *S. aureus* gene coding for β-lactamase in *B.subitilis*. The presence of the *bla* promoter, coding region for the signal sequence and first few a.a. of the native protein, destabilised a pC194 based shuttle vector which had previously been stable in *B. subtilis*. Therefore, it would appear that as well as the general stability problems encountered using recombinant plasmids in *B. subtilis*, the promoter regions of *S. aureus* genes may be incompatable with the *S. aureus/B. subtilis* shuttle vector systems currently used, by virtue of their ability to cause an overproduction of the heterologous protein to an extent which is deleterious to the host organism.

It has since been established that the problems of structural and segregational instability often observed when using *B. subtilis* as a recombinant host were largely due to the type of plasmid employed in vector construction. The plasmids most frequently used were the small, staphylococcal antibiotic resistance plasmids which replicate by a rolling circle mechanism. A characteristic feature of this mode of replication is that it is easily perturbed. It follows that alterations to plasmid integrity (i.e., during vector construction or insertion of cloned DNA), or the use of plasmids in an alternative host, frequently leads to structural or segregational instability (see Gruss and Ehrlich, 1989).

Use of an alternative plasmid which does not replicate by a rolling circle mechanism, e.g., pIP404 and pAMβ1 is now available, pAMβ1 which replicates using a unidirectional theta mechanism, has been experimentally shown to result in 1000-fold greater structural stability in *B. subtilis* over a ss DNA (pC194) control (Janniere *et al.*, 1990).

1.2.2 Gram-Negative Host

One of the major considerations when deciding on an alternative host system for expression of a particular gene is the probability that the transcription and translation regulation regions of the heterologous gene are recognised by the new host. This

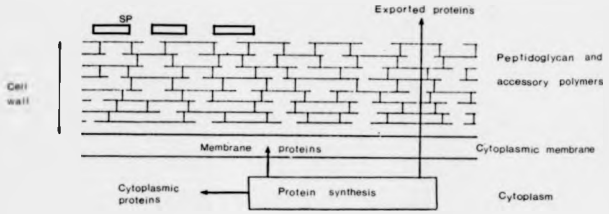
concern increases the higher the degree of diversity between the two systems. Thus, a gene from a Gram-positive organism is less likely to be correctly expressed in a Gram-negative host than in another Gram-positive host. In general, it has been shown that genes from Gram-positive bacteria are more likely to be successfully expressed in a Gram-negative bacterium than *vice versa*. It is thought that the RNA polymerase enzymes of Gram-positive bacteria are more stringent in their requirements for promoter recognition than those of *E. coli*. A barrier to gene expression also exists at the translational level. Thus Gram-positive genes, and in particular *B. subtilis*, have shown a requirement for a more extensive degree of homology between the mRNA Shine-Delgarno sequence and the 3′ end of the 16s rRNA than that required by *E. coli* (Murray and Rabinowitz, 1982). It is, however, known that the *spa* gene from *S. aureus* Cowan I is expressed in *E. coli* and produces a functional protein, as the initial detection of the cloned *spa* gene from an *E. coli* host relied upon the production of a protein exhibiting IgG binding activity (Duggleby and Jones, 1983). The levels of functional SpA produced in *E. coli* (0.2% total protein) were 10-fold less than those produced by the chromosomally located *spa* gene in *S. aureus* Cowan I, despite being located on pBR328 (a multicopy plasmid derived from pBR322). This may be caused by a number of factors, including the poor recognition of the *spa* transcription or translational regulatory sequences, or may result from the incorrect processing of the SpA polypeptide, yielding a high proportion of inactive SpA.

The use of *E. coli* for expression of heterologous proteins is popular because of the ease with which the genome can be manipulated. The wide range of *E. coli* host strains and plasmid vectors available allows a high degree of control over factors such as plasmid stability, the production of some proteases, plasmid copy number, promoter efficiency and regulation, translation initiation strength, selection of marker for plasmid propagation and maintenance, and a well characterised origin of replication. Transformation of *E. coli* is highly efficient and relatively simple to perform, while other laboratory techniques used in the manipulation of recombinant DNA in *E. coli* are well documented (Maniatis *et al.*, 1982, Rodriguez and Tait, 1983 and Walker, 1984).

Therefore, *E. coli* may prove a highly suitable host for the generation of SpA in order to study the properties of the gene, as it allows the easy manipulation of the various structural and regulatory regions of that gene.

A problem that is likely to be encountered in using *E. coli* as a host for SpA production, is the final cellular location of the heterologous protein. The cell wall of *E. coli*, a Gram-negative bacterium, is substantially different to that of a Gram-positive bacterium such as *S. aureus*, (Figure 1.2). In Gram-positive bacteria, proteins are located in the cytoplasm, in or on the cytoplasmic membrane or exported to the growth medium, whereas in Gram-negative bacteria, proteins can also be localised to the periplasmic space or to the inner or outer membrane. This difference in cell wall structure is reflected by the fact that *E. coli* in common with most enterobacteria excrete very few proteins to the exterior of the cell. Those that are fully exported are mainly toxins: (e.g., haemolysin, enterotoxins and bacteriocins such as colicin) and do not appear to rely on a common pathway for transport across the outer membrane. As is shown in the summary below, each of the above toxins utilises a different mode of exit from the cell. The heamolysin determinant consists of four genes, *hly*C, *hly*A, *hly*B and *hly*D of which only *hly*A codes for the heamolysin protein, with the export being dependent on the products of *hly*B and *hly*D. These helper proteins are believed to interact with the C-terminus of the *hly*A protein, allowing ATP-dependent transport through a transenvelope channel (Felmlee *et al.*, 1985a and b; Mackman *et al.*, 1985). There are two *E. coli* enterotoxins. The heat stable enterotoxin is a monomer, and secretion is thought to be dependent on the small size (5 kDa) of the molecule. The heat labile toxin is a multimeric complex composed of an A subunit and five B subunits. The A subunit requires the presence of the B subunit for release (Yamamoto and Yokota, 1982), which is thought to be facilitated by the conditions of low pH, low oxygen and bile salts present in the intestine (Hirst *et al.*, 1984). Colicins are encoded on naturally occurring plasmids, which generally encode a lysis protein required for colicin release.

a) Gram-positive cell envelope



b) Gram-negative cell envelope

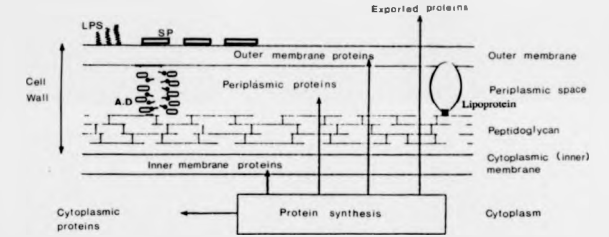


Figure 1.2 Schematic representation of the cell envelope of a typical Gram-negative and Gram-positive bacteria, showing the localisation of proteins

LPS: lipopolysaccharide chain. SP: surface proteins, consisting of additional surface layers such as protein matrices and capsules; and surface appendages such as fimbriae, pili and flagellae. AD: adhesion zone: an alternative route for secretion of proteins via a transenvelope structure rather than through the periplasm (Bayer, *et al.*, 1982).

This type of release is not considered to be true secretion as it is nonspecific and causes envelope damage by activating a phospholipase, resulting in the release of free fatty acids and alteration in outer membrane permeability (Pugsley and Schwartz, 1984).

In contrast to the above examples, transport across the *E. coli* cytoplasmic membrane of most *E. coli* proteins follows the "general" secretion pathway involving the *sec* genes and the presence of a signal sequence at the N-terminus of the polypeptide, as described in Section 3.3.3. In general terms the periplasmic export of proteins in *E. coli* and the secretion to the growth medium of proteins by Gram-positive bacteria are considered an almost identical process, requiring similar transportation signals. Therefore, any signal sequence present in the SpA polypeptide which would normally allow secretion to the cell wall or growth medium in *S. aureus*, even if recognised in *E. coli* may only allow the transport of the SpA polypeptide across the inner membrane to the periplasmic space. The final cellular location in *E. coli* of a heterologous protein normally found in the cell membrane or secreted to the growth medium of a Gram-positive organism appears to vary according to the individual protein expressed. Thus the α-amylase gene from *Bacillus coagulans* encodes for an exoprotein in its natural host, however, when expressed in *E. coli* the protein is secreted to the periplasmic space (Cornelis *et al.*, 1982). In contrast, Sarvas and Palva (1983) reported the export of the penicillinase encoded by the *Bacillus licheniformis* penicillinase (*pen*P) gene to the outer membrane fraction of *E. coli*. Here it is present as an amphiphilic protein with properties indistinguishable from those of the *B.licheniformis* membrane penicillinase, however, *E. coli* appears to lack the proteolytic enzymes necessary to cleave the membrane-bound lipoprotein form of the penicillinase to give the soluble exopencillinase found in the natural host.

The cellular location of a heterologous protein produced by *E. coli* can dramatically affect the final levels of protein recovered from a cell culture, both quantitatively and qualitatively. It has been shown to have a direct affect on the ratio of active to inactive forms of the protein, the degradation rate, the ease of final recovery of the product and

finally on the growth rate and viability of the host organism itself. A protein that is normally located outside of the cytoplasm in its natural host, if retained in the cytoplasm when expressed in a heterologous host such as *E. coli*, will often have little or none of the normal enzymatic activity. This is thought to be due to the incorrect folding of the protein brought about by the reducing environment of the cytoplasm. A high degradation rate is often seen for non-cytoplasmic proteins retained in the cytoplasm. This may be caused by a combination of two factors, the heterologous protein may not possess N- and C-terminal sequences resistant to the action of cytoplasmic proteases and the fact that there are a greater number of proteolytic enzymes present in the cytoplasm than in the rest of the cell. Eight separate soluble proteolytic acitvities have been reported in *E. coli* (Swamy and Goldberg, 1981). Five were located in the cytoplasm, two in the periplasm and one distributed evenly between the two compartments. Talmadge and Gilbert, (1982) have shown dissimilar rates of degradation for a hybrid proinsulin molecule when located in different cellular compartments. When secreted to the periplasm, the hybrid protein exhibited a 10-fold longer survival time than that remaining in the cytoplasm. Synthesis of high levels of some recombinant proteins in *E. coli* have resulted in the production of intracellular inclusion bodies (Klier *et al.*, 1982; Weis *et al.*, 1983; Williams *et al.*, 1982; Schoemaker, *et al.* 1985). The sequestering of the heterologous protein into dense insoluble inclusion bodies has been reported to protect a high percentage of the protein against proteolysis (Weis *et al.*, 1983; Kleid *et al.*, 1981; Cheng *et al.*, 1981). Inclusion bodies can be advantageous in increasing the level of recovery of the protein by making the isolation and purification stages easier. However although some inclusion bodies are easily solubilised to yield large amounts of conformationally correct protein that retain their enzymatic activity, it is more common for the inclusion bodies to be composed of aggregates of the protein which are very difficult to solubilise and purify. Schoemaker *et al.* (1985) showed that in the case of recombinant prochymosin, aggregation was caused by the interlinkage of intermolecular disulphide bonds to form protein multimers. Though disulphide bonds are usually unable to form in the

cytoplasm, it is suggested that they may be able to exist in an inclusion body, as the molecules are very closely packed and are therefore more likely to react with one another and as such are protected from the reducing environment of the cytoplasm. The presence of cysteine residues, and thus the potential for disulphide linkages, is a common feature in proteins seen to accumulate as inclusion bodies when present in high concentrations.

Secretion to the periplasm or outer membrane of *E. coli* would appear the preferable location for non-cytoplasmic heterologous proteins, in terms of reducing degradation and encouraging the correct post translational processing and final protein conformation while avoiding the formation of inclusion bodies. However, Emr *et al.* (1978; 1980) have reported that at high levels of protein production, secretion of the protein to the periplasm can be highly toxic to the cell, resulting in either reduced growth or cell lysis. In these studies cell lysis was brought about by the production of large amounts of a *lamB -lacZ* hybrid protein and the cells inability to efficiently export the hybrid protein to the periplasm, resulting in a lethal jamming of the normal cellular protein export machinery. Similar results were obtained by Rose and Shafferman (1981) when expressing the vesicular stomatitis glycoprotein. In this case cellular lethality was prevented by removing the hydrophobic core region of the signal sequence thereby almost certainly preventing its insertion into the cytoplasmic membrane. Brosius, (1984) revealed similar results when expressing a hybrid rat insulin gene, but also reported detrimental effects to the cell of a non-excreted form of the rat insulin gene when fused to an extremely strong promoter (rrnB). These latter observations indicate that other, more specific, effects may contribute to cellular lethality in individual gene cases. In such circumstances the formation of cytoplasmic inclusion bodies may serve to protect the host cell against the toxic effects of overproduction (Weis *et al.* 1983).

Despite there being problems of poor efficiency, poor yields and toxicity to cells when expressing a heterologous protein in *E. coli* these are only potential problems and might

not arise in the case of SpA. Even if one or more were found to be rate limiting factors, the versatility that is possible using different *E. coli* hosts and plasmid systems would offer a greater chance of dealing with such a problem, than if it was found to occur in the less well understood Gram-positive systems that are currently available. Therefore *E. coli* was selected as the preliminary host for the development of a high expression system for SpA.

1.3 AIMS AND RESEARCH OBJECTIVES

The primary aims of the project undertaken during the course of this PhD thesis was to determine the nucleotide sequence of the *S. aureus* Cowan I protein A gene and its surrounding regulatory regions. Once established it was anticipated that knowledge of the nucleotide sequence would facilitate the development of other areas of the project such as: i. a study of the efficiency of Gram-positive gene regulatory signals in *E. coli*; ii. the cloning of the *spa* gene, to enable the high expression of SpA; and SpA-fusion proteins in a non-pathogenic host for commercial exploitation; iii. the performance of theoretical studies on the predicted a.a sequences and secondary structure of the encoded protein.

CHAPTER TWO

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemicals and Materials

Except where stated below, all chemicals were purchased from BDH Chemicals Ltd., Poole, Dorset, and were of analytical reagent grade (AnalaR) wherever possible.

Amersham International, Bucks, England
[γ-^{32}P] dATP (50% ethanol), [α-^{35}S] thio-dATP, Na^{125}Iodine. Protein molecular weight size markers [α-^{32}P] dATP (aqueous), [^{14}C] methylated protein mixture.

Anderman & Co., East Molesley, Surrey, England.
Nitrocellulose sheets, nitrocellulose filter discs, manufactured by Schleicher & Schull.

Bethesda Research Laboratories (BRL/GIBCO), Paisley, Scotland
The following chemicals were 'ultra pure' grade where possible. Urea, caesium chloride, agarose, ammonium persulphate, formamide, isopropyl thio-β-galactoside (IPTG), phenol, tetramethylethylenediamine (TEMED), Lambda DNA, new born calf serum.

Biorad Laboratories, Caxton Way, Watford, Herts
Acrylamide, methylene-bis acrylamide.

Difco Laboratories, West Molesley, Surrey, England.
Tryptone, yeast extract, agar.

ICN Biomedicals Ltd, High Wycombe, Bucks
Gelbond

May and Baker Ltd.
Glycerol, diethyl ether.

Biocell Ltd, Cardiff, Wales
Goat anti-rabbit labelled with gold.

Pharmacia (LKB) Ltd, Milton Keynes, England.
Sephadex G25 columns (PD-10), Sephadex G50 Protein A.

Sigma Chemical Company (UK) Ltd., Poole, Dorset,England
Ethidium bromide, ampicillin, chloramphenicol, tetracycline, bromophenol Blue, xylene cyanol, orange G, ficoll, adenosine 5′-triphosphate (ATP), DL-dithiothreitol (DTT), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (BCIG/X-gal), isoamyl alcohol , dimethylformamide (DMF), polyethylene glycol (PEG; M_r 6000), sodium deoxycholate, polyvinylpyrrolidone, bovine serum albumin (BSA pentax fraction V), human IgG, salmon sperm DNA, thiamine, Triton X-100, Coomassie Brilliant Blue R250, polyoxyethylene 20 cetyl ether (Brij 58), nitrophenyl phosphate.

2.1.2 **Enzymes**

Boehringer Mannheim (BCL), Lewes, Sussex, U.K.
Calf intestinal alkaline phosphatase, T4 DNA polymerase, Klenow DNA polymerase I, Proteinase K, RNase T1, DNaseI.

Bethesda Research Laboratories (BRL/GIBCO), Paisley, U.K.
All restriction endonucleases, T4 polynucleotide ligase, T4 polynucleotide kinase.

Sigma Chemical Co.Ltd.,Poole,U.K.
Lysozyme (Grade I), Lysostaphin

2.1.3 **Media**

All media listed below were prepared in double distilled, de-ionised water and autoclaved at 15 p.s.i. (121°C) for 15 min unless otherwise stated.

2.1.3.1. Routine Growth Media

L-broth. This medium was routinely used for the cultivation of *E.coli* during the protein A project.

	gl^{-1}
Bacto-tryptone	10.0
Yeast extract	5.0
NaCl	5.0

The pH was adjusted to 7.4 using 1M NaOH. For solid media (L-agar) 2% (w/v/purified agar was added).

2xYT Broth: This media was routinely used for the cultivation of *E. coli* during the alkaline phosphatase project and for the growth of phage containing *E.coli* JM101.

	gl^{-1}
Bacto-tryptone	16.0
Bacto-yeast extract	10.0
NaCl	10.0

The pH was adjusted to 7.4 using 1 M NaOH. For solid media (2xYT agar), 2% (w/v) purified agar was added.

Low phosphate broth (LP)

	gl^{-1}
Tris base	14.54
NaCl	4.68
KCl	1.49
NH_4Cl	1.07
Na_2SO_4	0.97
$MgCl_2$	0.203
$CaCl_2$	0.0294
$ZnCl_2$	0.0003

The pH was adjusted to 7.5 with 1M NaOH. For solid media LPB-agar, 1.5% (w/v) purified agar was added prior to autoclaving. Before use, the following supplements were added:

	Final Concentration	
Bactopeptone	5%	autoclaved separately
Casamino acids	0.05%	autoclaved separately
Thiamine	20µg/ml	(filter sterilised)
Glucose	0.2%	(filter sterilised)
Ampicillin	50µg/ml	(filter sterilised)

H-Top Agar: This medium was used for the soft agar overlays in the cultivation of M13 phage plaques.

	gl^{-1}
NaCl	8.0
Agar	8.0

The pH was adjusted to 7.4 with 1 M NaOH.

M9 Salts (10x stock): This defined minimal medium was used for the cultivation of *E. coli* JM83 [SpA]

	gl^{-1}
Na_2HPO_4	300
KH_2PO_4	150
NaCl	25
NH_4Cl	150

The pH was adjusted to 7.4 with 10 M NaOH, and the mixture autoclaved. A volume of 20 ml of this 10x stock was added to 180 ml molten (50°C) agar/water 2% (w/v) prior to the addition of the following constituents:

1 ml 100 mM $MgSO_4$	(autoclaved separately)
1 ml 10 mM $CaCl_2$	(autoclaved separately)
2 ml 20% (w/v) Glucose	(filter sterilised)

2.1.3.2. Antibiotics

Antibiotics were incorporated at appropriate concentrations in both solid and liquid media for the selection of resistant bacterial clones.

Ampicillin

Stock solution: 10 mg ml^{-1} of the sodium salt of ampicillin in distilled water. This was sterilised by filtration (0.22 µm Millipore disposable filter) and stored at -20°C.

Working concentration: 50µg ml^{-1}.

Chloramphenicol

Stock solution: 5 mg ml^{-1} in 100% ethanol (v/v). This was stored at -20°C.

Working concentration for selection of resistant bacteria: 30 µg ml^{-1}, for amplification of plasmids: 170µg ml^{-1}.

Tetracycline

Stock solution 5 mg ml^{-1} of tetracycline hydrochloride in ethanol/water (50% v/v). This was stored at -20°C in the dark.

Working concentration: 12.5$\mu g ml^{-1}$.

2.1.4 Buffers and Solutions

All solutions listed below were prepared in deionised, double distilled water and autoclaved at 15p.s.i.(121°C) for 15 min unless otherwise stated.

TE Buffer (1x):
Tris-HCl 10 mM
Na_2EDTA 0.1 mM
pH 8.0

TES Buffer (1x):
Tris-HCl 10 mM
Na_2EDTA 0.1 mM
NaCl 170 mM
pH 8.0

SSC Buffer (20x):
NaCl 3.0 M
$Na_3Citrate$ 0.3 M
pH 7.0

TBE Buffer (10x):
Tris base 0.9 M
Boric acid 0.9 M
Na_2EDTA 0.03 M
pH 8.3

TM buffer:
Tris-HCl 0.1 M
$MgCl_2$ 0.05 M
pH 8.5

Phosphate Buffered Saline (PBS):
Na_2HPO_4 8 mM
KH_2PO_4 1.5 mM
NaCl 137 mM
KCl 2.7 mM
pH 7.4

Tricine buffer (4x):

	gl^{-1}
Tris base	38.8
Tricine	17.8
Calcium Lactate	0.5

Ligation buffer (10x):

Tris-HCl	0.3 M
NaCl	0.3 M
$MgCl_2$	0.075 M
Spermidine	0.01 M
ATP (Na salt)	0.0025 M
DTT	0.02 M
Na_2EDTA	0.002 M

pH 7.5 :Stored at -20°C.

Kinase buffer (10x):

Tris-HCl	0.5 M
$MgCl_2$	0.1 M
DTT	0.05 M

pH 7.4
Stored at -20°C

T4 polymerase buffer (10x)

Tris-acetate	0.33 M
K acetate	0.66 M
Mg acetate	0.10 M
DTT	0.005 M
BSA	1 mg ml^{-1}

(pentax Frac V)
pH 7.9
Stored at -20°C.

S.T.E.T buffer:

Sucrose	8 % (w/v)
Triton X-100	5 % (v/v)
Na_2EDTA	0.05 M
Tris-HCl	0.05 M

pH 8.0

Brij/Doc Solution

Brij 58	1%
Sodium deoxy-cholate (DOC)	0.4%
Tris	0.01 M
Na_2EDTA	0.001M

pH8.0

Birnboim lysis solution
(to make up 1ml)

Lysozyme	200 mg
Tris-HCl (pH8.0)	25 mM
EDTA (ph8.0)	10 mM
Glucose	50 mM

(this solution was prepared on day of use) and stored

Birnboim alkaline SDS Solution

NaOH	0.2 N
SDS	1%

RbCl Transformation buffer A

Mops pH7.0	10 mM
RbCl	10 mM

(this was sterilised by filtration using a 0.22 µm Millipore disposable filter) and stored at 4°C

RbCl Transformation buffer B

Mops pH6.5	100 mM
$CaCl_2$	50 mM
RbCl	10 mM

(this was sterilised by filtration using a 0.22 µm Millipore disposable filter) and stored at 4°C

Agarose gel (Ficoll) tracking dye

Ficol	50% w/v
Bromophenol	0.02% w/v
Xylene cyanol	0.02% w/v
Orange G	0.08% w/v

Denhardts reagent (50x)

	$g\ l^{-1}$
Polyvinylpyrrolidone	10
BSA (Pentax Frac V)	10
Ficoll	10

Filter sterilised and stored at -20°C.

Prehybridisation solution

SSC	6x
Denhardts reagent	10x
SDS	0.2% (w/v)
Denatured SS salmon sperm DNA	100 $\mu g ml^{-1}$

The salmon sperm DNA was denatured by heating to 100°C for 5 minutes.

40% (w/v) Acrylamide stock

	$g\ l^{-1}$
Acrylamide	380
Bis-acrylamide	20

Made up to 1 l with distilled water,de-ionised with Amberlite MB-1 resin, and filtered.

0.5x TBE gel mix:

40% (w/v) acrylamide	150 ml
10x TBE	50 ml
Urea (ultra-pure)	460 g

Made up to 1 l with distilled water.

5.0x TBE gel mix:

40% acrylamide	150 ml
10x TBE	500 ml
Urea (ultra-pure)	460 g
Bromophenol blue	0.05 g

Made up to 1 l with distilled water.

Formamide gel loading dye:

Deionised formamide	95% (v/v)
Xylene cyanol FF	0.05% (w/v)
Bromophenol blue	0.05% (w/v)
Na_2EDTA	20 mM

SDS acrylamide stock solution:

	$g\ l^{-1}$
Acrylamide	300
DATD	8

Made up to 1 l with distilled water, de-ionised with Amberlite MB-1 resin, and filtered.

SDS-PAGE resolving buffer:

Tris-HCl	0.75 M
SDS	0.2% (w/v)
pH 8.8	

SDS-PAGE stacking gel buffer

Tris-HCl	0.25 M
SDS	0.2% (w/v)
pH 6.8	

SDS-PAGE electrophoresis buffer

Tris	25 mM
Glycine	190 mM
SDS	0.1% (w/v)
pH 8.3	

SDS-PAGE sample loading buffer

Tris-HCl	0.2M
Glycerol	10% (v/v)
β-mercaptoethanol	10% (v/v)
SDS	10% (v/v)
Bromophenol blue	0.005%(w/v)
pH 6.8	

Monorocket gel staining solution

Coomassie brilliant blue R250	0.45%(w/v)
Glacial acetic acid	100 ml
Methanol	450 ml

Monorocket gel destaining solution

Glacial acetic acid	10% (v/v)
Methanol	45% (v/v)

Western blot transfer buffer

Tris	25 mM
Glycine	192 mM
Methanol	20% (v/v)
pH 8.3	

GAPDH Assay Mix

	In total vol of 30 ml
50mM Sodium pyrophosphate, pH 8.5	27 ml
*Fructose 1-6 diphosphate (85 mg ml^{-1})	1.0 ml
*NAD (10 mg ml^{-1})	0.5 ml
0.3 M Sodium arsenate	0.5 ml

*make up fresh

2.1.4.1 Stocks and Solutions for dideoxy-sequencing method used in Section 2.2.10.4

10x annealing buffer
100 mM Tris, pH 8.5
100 mM $MgCl_2$

dNTP - mix buffer
50 mM Tris, pH 8.0
1 mM EDTA

Dideoxytriphosphates (ddNTP) solutions
Stocks of 10 mM dideoxytriphosphates were kept frozen (-20°C) in 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA). Working stocks were diluted to the follow level. 0.3 mM ddTTP, 0.1 mM ddCTP, 0.15 mM ddGTP, 0.03 mM ddATP.

Deoxytriphosphate (dNTP) solutions
Stocks of 20 mM dNTPs and 0.5 mM working solutions were kept frozen (-20°C) in water.

Deoxytriphosphate (dNTP°) mixes

	T° dTTP	C° dCTP	G° dGTP	A° dATP
0.5 mM dATP	1 µl	20 µl	20 µl	20 µl
0.5 mM dCTP	20 µl	1 µl	20 µl	20 µl
0.5 mM dGTP	20 µl	20 µl	1 µl	20 µl
50 mM Tris, pH 8.0, 1 mM EDTA	5 µl	5 µl	5 µl	5 µl

the dNTP° were kept frozen (-20°C)

200 µl of 0.5 mM ATP in water was used for the chase.

2.1.4.2. Stocks and Solutions for the dideoxy-sequencing Method Used in Section 2.2.10.5.

Dideoxytriphosphate (ddNTP) solutions
Working stocks of 10 mM dideoxyphosphate were kept frozen (-20°C) in T.E. buffer.

Deoxytriphosphate (dNTP) solutions
Stocks of 50 mM deoxytriphosphate were kept frozen (-20°C) In T.E. buffer and working solutions at 0.5 mM in T.E. buffer.

Dideoxytriphosphate/Deoxtriphosphate mix (NTP)
These mixes were made from the working solutions as follows:

	"T"	"C"	"G"	"A"
0.5 mM dTTP	25	500	500	500
0.5 mM dCTP	500	25	500	500
0.5 mM dGTP	500	500	25	500
10.0 mM ddTTP	50			
10.0 mM ddCTP		8		
10.0 mM ddGTP			16	
10.0 mM ddATP				2
T.E. buffer	1000	1000	1000	500

dNTP chase mix was made up of 0.25 mM of all 4 dNTP stocks diluted with T.E. buffer.

2.1.5 Bacterial Strains

The bacterial strains used in this study are listed below.

STRAIN	DESCRIPTION/GENOTYPE	SOURCE
E. coli MC1061	*hsdR*,*mcr*B, *ara*D139, (*ara*ABC-*leu*) *lac*X74,*gal*U,*rsp*L,*thi*	N. Minton
E.coli HB101	*sup*E44,*hsd*S20 ($r_B^- m_B^-$)*rec*A13, *ara*-14,*pro*Az,*Lac*Y1,*gal*K2,*rps*L20,*xyl*-5, *mtl*-1	S. Jones
E.coli JM83	*ara* (*lac-pro*AB), *str*A, *thi* (80d*lac*I^q, Z MI)	N. Minton
E. coli JM101	*sup*E,*thi* (*lac-pro*AB) F'{*tra*D36, *pro*AB$^+$, *lacI*^{q},*lacZ* M15]	N. Minton
E.coli JLM85	Derivative of JM83 carrying the *rec*A from JA221	N.Minton
E. coli TG2	*sup*E, *hsd*S, *thi* (*lac-pro*AB) [F',*tra*D,*pro*AB$^+$,*lacI*^{q},*lacZ*,M15]*rec*A	N. Minton
E.coli GM242	*dam*$^-$3,*dcm*6,*thr*,$^-$*leu*$^-$,*thi*$^-$, *lac*Y *gal*K2,*gal*T22,*ara*14,*ton*A31,*tsx*78, *sup*E44	A. Akrigg
E.coli AW1046	*pho*A, *ara*D139, (*ara*, *leu*)7697, *lac*X74,*gal*U, *gal*K, *str*A.	N. Minton
*E. coli*BMH71-18mutL	(*lac*,*pro*),*thi*[F'*pro*AB$^+$, *lacI*|*lacZ*,M15],*sup*E	D.A. Barstow
S. aureus Cowan I	(NCTC8530)	C. Duggleby
S. aureus Wood 46	(NCTC7121)	R. Sharp

2.1.6 Plasmids and Phages

The plasmids and phages used in this study were available in the Division and are shown below.

PLASMID	MARKER	REFERENCE
pBR327	Ap^r	Soberon *et al* 1980
pUC9	Ap^r	Vieira and Messing, 1982
pUC8	Ap^r	Vieira and Messing,1982
pNM52	Ap^r	Gilbert *et al*, 1986
pMTL20	Ap^r	Chambers *et al*, 1988

PHAGES	REFERENCE
M13mp8	Messing and Vieira,1982
M13mp9	Messing and Vieira,1982

2.2 METHODS

2.2.1. Growth and Preservation of Bacterial Strains and Phages

2.2.1.1 *E. coli*

E. coli strains were routinely grown aerobically at 37°C in either 2xYT broth or L-broth. These strains were also cultivated on 2% agar solidified 2xYT broth or L-broth. Where necessary the medium was supplemented with the appropriate antibiotic. When required the plate medium was supplemented with X-gal (40 mg ml^{-1} in dimethylformamide) at 0.08 mg ml^{-1} and with IPTG (20 mg ml^{-1} in distilled water) at 0.05 mg ml^{-1}. Cultures were maintained for the short term at 4°C on L-agar plates. Longer term storage was at -70°C as broth cultures supplemented with 15% (v/v) glycerol.

2.2.1.2 Bacteriophage M13

Bacteriophage M13 was propagated in the host strain BMH71-18 *mut*L and grown in 2xYT broth. For preparation of plaques, phage was added at the appropriate dilution to 3 ml H-top agar at 47°C containing 200 µl of an exponential culture or BMH 71-18. When required the H-top agar was supplemented with 30 µl X-gal (40 mg ml^{-1} in dimethyl formamide) and 20 µl IPTG (20 mg ml^{-1} in distilled water). The H-top agar was poured over solid 2xYT medium and allowed to set prior to incubation at 37°C for 16 hours. Phage was stored as an infected *E. coli* culture at -70°C in 15% (v/v) glycerol.

2.2.2 Transformation and Transfection of *E. coli*

Strains of *E. coli* were transformed and transfected (with plasmid and M13 RF DNA respectively) by the rubidium chloride (RbCl) method of Kushner (1978).

2.2.2.1 Preparation of Competent Cells

The required *E. coli* strain grown overnight in L-broth at 37°C was used to inoculate 50 to 100 ml of prewarmed L-broth at a dilution of 1:50 in a 250 ml conical flask. The

culture was incubated at 37°C with vigorous aeration and the OD_{650} of the culture monitored at 15 to 30 min intervals. When the OD_{650} reached 0.16 to 0.18 the flask was placed on ice and 1.5 ml aliquots of the culture pipetted into chilled Eppendorf tubes, these were centrifuged at 13,000 rpm for 5 min at 4°C. The bacterial pellet was washed in 1 ml of chilled (4°C) transformation buffer A, centrifuged at 13,000 rpm at 4°C for 5 min and the pellet resuspended in 1 ml of chilled (4°C) transformation buffer B. After maintaining on ice for 30 min the cells were then centrifuged at 13,000 rpm at 4°C for 5 min, resuspended in 0.2 ml of chilled (4°C) transformation buffer B and maintained on ice prior to use.

2.2.2.2 Transformation with Plasmid and M13 RF DNA

Transformation was achieved by incubation of the transforming DNA (typically 20 to 200 ng of DNA in a maximum volume of 10 µl) with the 0.2 ml aliquots of the competent cells (2.2.2.1) on ice for 30 min, followed by a 30 sec heat shock at 45°C. Transformed cells were brought to room temperature, diluted in 0.5 ml of L-broth, incubated without shaking at 37°C for 1 hour (if antibiotic resistance required) then used to inoculate appropriate L-agar plates in the case of plasmid transformation, or diluted in 2xYT broth and used immediately to inoculate soft agar overlays in the case of M13 RF DNA transfections (2.2.1.2).

2.2.3 Treatment of DNA with Enzymes

2.2.3.1 Restriction Endonucleases

The reaction conditions were in accordance with manufacturers' instructions, except that a two to tenfold excess of enzyme was used routinely. The digestion products were analysed directly by agarose gel electrophoresis (2.2.6). If the restricted DNA was to be ligated the remaining restriction enzyme activity was destroyed by either heat inactivation at 70°C for 10 min (for heat labile enzymes), or by phenol extraction (2.2.4.7.1.).

2.2.3.2 T4 Polynucleotide Ligase

Ligation was typically carried out using DNA fragments at a final concentration of 15 to 30 $\mu g\ ml^{-1}$ in 1x 'ligation buffer' (2.1.4; Lathe *et al.*, 1984) which was added as a tenfold concentrate to the reaction mixture. T4 DNA ligase was added at a final concentration of 0.1 to 1.0 units μg^{-1} of DNA and the reaction mixture, in a 0.5 ml Eppendorf tube, incubated overnight at 15°C.

2.2.3.3 T4 DNA Polymerase

T4 DNA polymerase was used to convert 'sticky-ended' DNA to 'blunt-ended' DNA, in particular DNA fragments possessing recessed 5′ termini, by the method of O'Farrell, (1981).

The DNA sample was diluted in an appropriate volume (20 to 50 µl) of 1x T4 polymerase buffer supplemented with a one-tenth volume of dNTPs (0.25 mM each) to which 1 µl of T4 DNA polymerase was added. The reaction mixture was incubated at 37°C for 30 min and the enzyme then inactivated by heat treatment (70°C for 10 min). Alternatively, the treated DNA was then purified by phenol extraction (2.2.4.7.1) and ethanol precipitation (2.2.4.7.3).

2.2.3.4 DNA Polymerase, Large Fragment (Klenow Fragment)

Klenow polymerase was used to "blunt end" recessed 3′ termini by the addition of sequential bases (i.e end-labelling of *Hin*dIII digested Lambda DNA). All such reactions were carried out in 1xTM buffer (2.1.4) supplemented with a one-tenth volume of 0.25 M dNTPs, and 1 ml Klenow polymerase. Reactions were carried out at room temperature for 30 min. Remaining polymerase activity was inactivated by heat treatment (70°C for 10 min).

2.2.3.5 Calf Intestinal Phosphatase (CIP)

Dephosphorylation of the 5′ terminus of DNA was carried out using calf intestinal phosphatase (CIP). Following digestion of the DNA with the appropriate restriction

endonuclease, CIP was generally added directly to the reaction tube, at a ratio of 1 to 5 units of CIP μg^{-1} of DNA, and the tube incubated for a further 30 to 60 min at 37°C. Remaining CIP was removed by phenol extraction (2.2.4.7.1) and subsequent ethanol precipitation (2.2.4.7.3).

2.2.3.6 T4 Polynucleotide Kinase

T4 polynucleotide kinase was used to phosphorylate the 5'-hydroxyl terminus of synthetic oligonucleotides for use as primers for DNA sequencing, site directed mutagenesis or as radio-labelled oligonucleotide probes.

2.2.3.7 DNase free RNase

RNA was removed from DNA samples using DNase free RNase. Stock solutions of RNase A (10 mg ml^{-1} in 10 mM Tris-HCl, pH 7.5) were boiled (100°C) for 15 min prior to use or storage at -20°C.

DNA samples were treated with RNase at a final concentration of 25-50 $\mu g\ ml^{-1}$ by incubation at 37°C for 1 hour.

2.2.3.8 Determination of Conditions for Partial Digestion of Chromosome DNA with Restriction Endonucleases

This method was used to obtain a partial digestion of *E. coli* JM83 genomic DNA with *Sau*3A.

A reaction mixture was prepared which contained 50 µg of *E. coli* genomic DNA in the appropriate restriction enzyme buffer in a final volume of 150 µl. This mixture was aliquoted into eppendorf tubes as follows; 30 µl into tube 1, 15 µl into tubes 2 to 8, and the remainder into tube 9. All the tubes were chilled on ice for 10 min and then 20 units of restriction enzyme was added to tube 1 and gently mixed. This gave an enzyme concentration of 2 units/µg DNA. Half of this (15 µl) was transferred to tube 2, and this step was repeated to give serial dilutions from tube 1 to 8. None of the restriction

enzyme mix was added to tube 9. Tubes 1 to 8 were then incubated at 37°C for 60 min and the reactions stopped by chilling on ice and the addition of 1 µl of 0.5M Na_2 EDTA. A volume of 3 µl of tracking dye was then added to all nine tubes and the samples electrophoresed on a 0.8% agarose gel. Subsequent staining of the gel revealed the appropriate restriction enzyme concentration required to achieve partial digestion of the DNA.

2.2.4 Isolation of DNA

2.2.4.1 Chromosomal DNA Isolation

The isolation of genomic DNA from *E. coli* JM83 was performed by a slight variation of the method of Marmur (1961).

E. coli JM83 was grown overnight in 1 litre of L-broth. The cells were harvested by centrifugation (Sorval RC5B, GSA rotor, 6,000 r.p.m., 10 min) washed in saline-EDTA, recentrifuged, and the wet weight of the harvest determined. The cells were resuspended in saline-EDTA at a density equivalent to 10 ml of saline-EDTA per gram of cells (wet weight). Lysozyme solution (10 mg ml^{-1} in distilled water) was added to the cell suspension at a dilution of 1 ml per 25 ml of cells and the mixture incubated at 37°C for 30 min. Hot SDS (60°C, 25% w/v) was added to give a final concentration of 2% and the mixture incubated at 60°C for 115 min with intermittent shaking. The mixture was then cooled to ambient temperature and 5 M sodium perchlorate added to give a final concentration of 1 M. An equal volume of chloroform : isoamylalcohol (24:1 v/v) was added and the emulsion gently shaken in a 10°C bath for 30 min. The emulsion was separated by centrifugation (Sorval RC5B, GSA rotor, 10,000 r.p.m., 5 min, 10°C). The aqueous phase (top layer) was carefully removed and transferred to a chilled (-20°C) glass measuring cylinder. Two volumes of chilled (-20°C) absolute ethanol were gently added and the DNA precipitation at the ethanol/aqueous interface, spooled onto a sterile glass rod, and left to drain. The DNA was resuspended in 75% of the original volume of 0.1 x SSC, and the SSC concentration adjusted to 1 x SSC with 10 x SSC. An equal volume of chloroform : isoamylalcohol (24 : 1 v/v) was again

added, the solution extracted, the DNA precipitated and spooled on to a glass rod, exactly as before. The DNA was again resuspended in 0.1 x SSC, this time in a volume equivalent to 50% of the original volume, and the SSC concentration adjusted to 1 x SSC with 10 x SSC. DNase-free RNase (2.2.3.7) was then added to a final concentration of 50 $\mu g\ ml^{-1}$ and the solution incubated for 1 hour at 37°C. SDS (25% w/v) was added to a final concentration of 0.5%, followed by the addition of proteinase K (20 $mg\ ml^{-1}$ in distilled water) to a final concentration of 1 $mg\ ml^{-1}$. The solution was incubated at 37°C for 60 min and then extracted three times with an equal volume of chloroform : isoamylalcohol (24 : 1 v/v). A 1/2 volume of phenol (2.2.4.7.1) was then added to the DNA solution and the mixture gently shaken at ambient temperature for 10 min. After separation of the phenol and aqueous phase by centrifugation (Sorval RC5B, SS34 rotor, 10,000 r.p.m., 5 min) the remaining traces of phenol were removed by 3 extractions with an equal volume of diethyl-ether (saturated with water). The aqueous phase was recovered and the DNA again precipitated with two volumes of chilled (-20°C) absolute ethanol, the DNA spooled onto a glass rod and the DNA eventually resuspended in a volume of 1 x SSC equivalent to 25% of the original volume. The DNA was finally precipitated by the addition of 1/10th volume of Acetate-EDTA (3.0 M sodium acetate; 0.001 M Na_2 EDTA; pH 7.0) and the gradual addition of a 0.54 volume of isopropanol whilst simultaneously spooling the DNA on to a sterile glass rod. The spooled DNA was then washed in 70%, 80% 90% and finally 100% ethanol, dried under vacuum, and resuspended in a minimum volume of 1 x TE buffer. The DNA concentration was determined spectrophotometrically by reading the optical density of a known dilution of the solution at 260 nm and the DNA solution stored at -20°C.

2.2.4.2 Small Scale Isolation of Plasmid DNA from *E. coli*

All small scale plasmid isolations were carried out using a modification of the rapid boiling method of Holmes and Quigley (1981), referred to as the STET method throughout the thesis; or (if a higher quality of DNA was required) using a modification

of the alkaline SDS method of (Birnboim and Doly, 1979) referred to as the mini-Birnboim method throughout this thesis.

2.2.4.2.1 STET method

A 1.5 ml sample of an overnight L-broth culture was centrifuged in an Eppendorf microcentrifuge (MSE Microcentaur, 13000 rpm for 30 sec). The bacterial pellet was resuspended in 350 µl S.T.E.T. buffer (2.1.4) by vortexing and 25 µl of a freshly prepared solution of lysozyme (10 mg ml^{-1}) in 25 mM Tris-HCl,pH 8.0 was added. After brief vortexing the suspension was placed in a boiling water bath for 40 sec. The suspension was immediately centrifuged (MSE Microcentaur, 13000 rpm for 10 min) the supernatant removed and reserved, and the pellet discarded. To the supernatant, 40 µl of 2.5M sodium acetate,pH 5.2 and 420 µl of isopropranol were then added, the tube contents briefly vortexed, and placed at room temperature for 5 min. The plasmid DNA present was recovered from the precipitate by centrifugation (MSE Microcentaur, 13000 rpm for 5 min at 4°C), the supernatant carefully decanted away, and the DNA pellet immediately washed with 1 ml 70% (v/v) ethanol, then dried under vacuum. The dried DNA pellet was resuspended in 34 µl 1xTE buffer and 1µl of DNase free RNase added (2.2.3.7). A 5 µl aliquot of this sample was sufficient for visualisation by agarose gel electrophoresis (2.2.6.1) following restriction endonuclease digestion (2.2.3.1).

2.2.4.2.2 Mini-Birnboim Method

A 1.5 ml sample of an overnight L-broth culture was centrifuged in an Eppendorf microcentrifuge (MSE Microcentaur, 13000 rpm for 80 seconds). The cells were resuspended in 100 µl of Birnboim lysis solution (2.1.4) by vortexing gently and chilled (0°C) on ice for 30 min. 200 µl of Birnboim alkaline SDS solution (2.1.4) was added, the tube contents briefly vortexed and chilled on ice for a further 5 min, after which 150 µl of 3M sodium acetate, pH 4.8 was added. The tube contents were gently mixed by inverting the tube 3 to 4 times, and placed on ice for 60 min, mixing occasionally during standing. The suspension was then centrifuged (MSE Microcentaur, 13000 rpm for 5 min) and 400 µl of the supernatant removed. 1 ml of ethanol was added to the

supernatant, it was mixed by inversion and then placed at -20°C for 30 min. The DNA pellet was recovered from the precipitate by centrifugation (MSE Microcentaur, 13000 rpm for 2 min), and the supernatant carefully aspirated using a fine drawn-out Pasteur pipette. The DNA pellet was resuspended in 200 µl of cold ethanol, placed at -20°C for 10 min, then recovered from the precipitate by centrifugation (MSE Microcentaur, 13000 rpm for 2 min), the supernatant carefully decanted away and the pellet was dried under vacuum. The dried pellet was resuspended in 75 µl of 1xTE buffer and 1 µl DNAse free RNAse (2.2.3.7) was added. 10 µl of this sample was sufficient for visualisation by gel electrophoresis (2.2.6.1) following restriction endonuclease digestion (2.2.3).

2.2.4.3 Large Scale Isolation of Plasmid DNA From *E. coli*

Plasmid DNA was purified from *E. coli* strains by the method of Clewell and Helsinki (1969).

500 ml L-broth supplemented with appropriate antibiotics was inoculated with 10 ml of an overnight culture and incubated overnight at 37°C with moderate aeration (200 r.p.m. on a shaker platform). The cells were harvested by centrifugation (Sorval RC5B, GSA rotor, 6000 r.p.m. for 10 min) and resuspended in 2 ml of 25% (w/v) sucrose, 50mM Tris-HCl,pH 8.0. Lysozyme, 0.3 ml of a 20 mg ml^{-1} solution in 25 mM Tris-HCl,pH 8.0, was added and the mixture left on ice for 3-5 min, after which time 2 ml of 250 mM Na_2EDTA was added. After a further 5 min on ice, 3 ml of Brij/Doc lysing solution (2.1.4) was added, the solution mixed gently. and left on ice for 20 min or until lysis occurred. The cell debris and the bulk of the chromosomal DNA was then pelleted by centrifugation (Sorval RC5B, SS34 rotor, 6000 r.p.m. for 30 min at 4°C) and the cleared lysate recovered by decanting the supernatant into a fresh tube. The supernatant was then subjected to isopycnic centrifugation (2.2.4.5.).

2.2.4.4 Large Scale Isolation of M13 (RF) DNA

Competent *E. coli* were transfected with single stranded M13 template DNA (2.2.2.2). Immediately following the 42°C heat shock the transfected cells were used to inoculate 10ml 2xYT broth which was incubated overnight at 37°C. The supernatant ("phage inoculum") was recovered following centrifugation (Sorval RC5B, SS34 rotor, 6000 r.p.m. for 10 min) and stored temporarily at 4°C. Meanwhile, 10 ml of an overnight culture of *E. coli* was used to inoculate 1 litre of prewarmed 2xYT broth. This culture was incubated at 37°C with vigorous aeration until an OD_{600} of 0.6 was reached. The 10 ml 'phage inoculum' was then added and the culture incubated for a further 6 hours at 37°C with vigorous aeration. The cells were harvested by centrifugation (Sorval RC5B, GSA rotor, 6000 r.p.m. for 10 min) and a cleared lysate prepared as described in section 2.2.4.3. The cleared lysate was then subjected to isopycnic centrifugation (2.2.4.5.)

2.2.4.5 Isopycnic Centrifugation of Plasmid and M13 (RF) DNA

Chromosomal and plasmid DNA were separated by centrifugation to equilibrium in caesium chloride-ethidium bromide density gradients (isopycnic centrifugation) as follows.

To each 1 ml volume of cleared lysate 0.95 g of solid caesium chloride and 0.02 ml ethidium bromide solution (10 mg ml^{-1} in distilled water) were added and mixed gently by inversion. The contents were transferred to a Du Pont/Sorval heat seal tube and the tube completely filled and sealed by heating with a Du Pont/Sorval tube heating system. If insufficient gradient mix was available the heat seal tubes were topped up with paraffin 'oil and balanced appropriately. The tubes were centrifuged in a Sorval Ti50 fixed angle rotor at 48000 r.p.m. (10 ml tubes) for 36 to 48 hours (20°C). After centrifugation, the plasmid or M13 RF (lower) and chromosomal (upper) DNA bands were visualised with a UV long wave transilluminator (Blak-ray, Ultra-violet products Inc., San Gabriel, Ca, U.S.A.). The lower band was removed by piercing the side of the tube with an 18-gauge hypodermic needle 0.5 cm beneath the band and withdrawing

it into a 2 ml syringe; an additional needle was used to pierce the top of the tube thereby facilitating a smooth removal of the desired band. Ethidium bromide was removed by extracting the sample 3 to 4 times with isoamyl alcohol which had been equilibrated with caesium chloride. The sample was then transferred to prepared 1/4" dialysis tubing and the caesium chloride removed by dialysis against one change of 1xTES buffer at 4°C followed by three changes of 1x TE buffer at 4°C. At least 1,000 volumes of dialysis buffer were used and the buffer changed after 1, 1, 2 and 4 hours of dialysis.

2.2.4.6. Recovery of DNA from Agarose Gels by Electro-Elution

DNA was extracted from agarose gels essentially as described by McDonnell *et al.*, (1977).

Following resolution of the desired DNA fragment by agarose gel electrophoresis and staining with ethidium bromide, the band was visualised with the aid of a long wave UV transilluminator (Blak-ray, Ultra-Violet Products Inc., San Gabriel, Ca, U.S.A.). The desired DNA band was excised from the gel with a sterile scalpel, placed in an appropriate length of 1/4" prepared dialysis tubing and submerged in 0.2xTBE buffer. The ends of the tubing were sealed and the dialysis bag placed lengthways at right angles to the current in a horizontal gel tank partially filled with 0.2x TBE buffer. Electroelution was carried out at 240 V for 2 hours or at 40 V overnight. After this time the current was reversed for several sec to disengage the eluted DNA from the dialysis tubing. The TBE buffer, containing the eluted DNA, was removed from the tubing and extracted once with phenol (2.2.4.7.1), three times with water saturated diethyl-ether (2.2.4.7.2), ethanol precipitated (2.2.4.7.3), and resuspended in an appropriate volume of TE buffer.

2.2.4.7 Solvent Extraction and Ethanol Precipitation of DNA

2.2.4.7.1 Phenol Extraction of DNA

Protein was removed from DNA samples by extraction with phenol, pre-equilibrated with TE buffer. A half volume of this TE-saturated phenol was mixed with the aqueous DNA sample and vortexed for 30 sec, left to stand for 2 min at room temperature, then vortexed again for 30 sec. After standing for 5 min the emulsion was separated by centrifugation (MSE Microcentaur, 13000 r.p.m. for 1 min). The aqueous phase was removed and transferred to a fresh tube. Diethyl-ether extraction was used to remove traces of phenol (2.2.4.7.2). The purified DNA sample was then ethanol precipitated (2.2.4.7.3).

2.2.4.7.2 Diethyl-ether extraction

Two volumes of diethyl-ether was added to the DNA sample, the tube vortex mixed for 30 sec and the layers allowed to separate. The diethyl-ether (top layer) and any material at the interface were removed and discarded. Diethyl-ether extractions were repeated until all traces of phenol had been removed (typically 2-4 times).

2.2.4.7.3 Ethanol precipitation

DNA was precipitated from solution by the addition of a one-tenth volume of 3M sodium acetate (pH 4.8) and two volumes of chilled absolute ethanol (-20^{o}C), followed by either incubation at -20°C for at least 1 hour or, incubation was at -70^{o}C for 15 min. After precipitation, the DNA was recovered by centrifugation (MSE Microcentaur, 13000 r.p.m. for 10 min). The supernatant was carefully decanted and discarded. The DNA pellet was washed in an appropriate volume of 70% (v/v) ethanol, prior to vacuum drying. The dessicated DNA pellet was resuspended in an appropriate volume of 1x TE buffer.

2.2.4.7.4 Isopropanol Precipitation

Rapid precipitation of plasmid DNA from a mixed solution of DNA and other materials was performed by precipitation of the sample with a 0.4 volume of 5M ammonium

acetate and two volumes of isopropanol, followed by incubation at room temperature for 10 minutes. The DNA was recovered by centrifugation (MSE Microcentaur, 13000 r.p.m. for 10 min) the pellet washed, in 70% (v/v) ethanol and the centrifugation step repeated. The DNA was then vacuum dried and resuspended in an appropriate volume of 1x TE buffer.

2.2.5 Spectrophotometric Analysis of DNA Preparations

2.2.5.1. Quantitative Analysis

A Pye Unicam SP6 spectrophotometer was used to quantitate the concentration of DNA in a given sample by measuring the absorbance at 260nm using quartz cuvettes. An OD_{260} reading of 1.0 was given by a concentration of 50μg ml^{-1} of double stranded DNA, 40μg ml^{-1} of single stranded DNA or 20μg ml^{-1} of synthetic oligonucleotide (Maniatis *et al.*,1982).

2.2.5.2 Qualitative Analysis

The degree of protein or phenol contamination in a DNA preparation, was estimated by determining the ratio of sample readings taken at 260nm and 280nm, where an $OD_{260/280}$ ratio less than 1.70, indicated significant protein contamination. Such samples were cleaned by additional phenol and diethyl ether extractions followed by an ethanol precipitation step (2.2.4.7.3).

2.2.6. Electrophoresis of DNA

2.2.6.1. Agarose Gel Electrophoresis

This was performed with horizontal slab gel electrophoresis tank systems. The gel systems employed were of the Schaffner type (Maniatis *et al.*, 1982), in which the gel is poured into a portable gel casting tray (20 x 20 cm gels), or of the 'mini-gel' type where the gel is cast directly into the electrophoresis apparatus (5 x 8 cm). The apparatus was obtained from BRL (H4 system) and Cambridge Life Sciences

respectively. The gel routinely used consisted of agarose (BRL ultra-pure) at concentrations between 0.8 and 1.5% (w/v). The gel was formed by boiling together (in a microwave oven) appropriate amounts of agarose, 1x TBE buffer and distilled water. The molten gel was allowed to cool to 60°C before pouring; 35 ml was routinely used for the horizontal mini gel and 250 ml was used for the analytical/preparative (20 x 20 cm) horizontal gel. Perspex combs of varying teeth size were inserted into the molten gels, prior to setting, to create the loading wells. Samples to be electrophoresed were mixed with a one-fifth volume of Ficoll tracking dye and loaded into the wells of the gel submersed in 1x TBE gel running buffer. The gel was then subjected to electrophoresis to give a voltage across the gel of between 1 and 5 V cm^{-1}. After electrophoresis the DNA was stained with 0.1 ml of ethidium bromide (10 mg ml^{-1} in distilled water) and visualised by medium wave UV illumination and photographed with Polaroid 667 black and white film (ASA 3000) using a red filter. Exposure times of between 1 and 4 sec at an aperture setting of F8 was usually sufficient.

2.2.6.2 Denaturing Polyacrylamide Gel Electrophoresis

DNA sequence reaction products were resolved by denaturing polyacrylamide electrophoresis using the thin gel system of Sanger and Coulson (1978). The gel plates (20 x 40 cm), one of which had 'rabbit ears' (the front plate) were cleaned thoroughly with pyroneg in solution followed by absolute ethanol. The front plate was then treated with 2% dimethyldichlorosilane in 1,1,1- trichloroethane (v/v). The back plate was treated with a solution comprised of 750µl 10% glacial acetic acid (v/v) and 75µl silane (τ-[methylacryloxy]-propyltrimethoxysilane) in 25 ml absolute ethanol. After allowing the plates to dry, they were lightly polished with absolute ethanol and the with a dry soft tissue to remove excess coating. The gel 'sandwiches' were then assembled with 0.35mm plastic card spacers (Raven Scientific Ltd) using 'nitto' tape (BRL). The 'gradient' gels routinely used were made from 40 ml 0.5x TBE gel mix and 7 ml 5x TBE gel mix. The gel mixes were polymerised by the addition of 25% (w/v) ammonium persulphate and TEMED to 0.2% (v/v) each. The gradients were poured in the following manner. Using a 10 ml pipette, 4 ml of the clear 0.5x TBE gel mix was

taken up, followed by the 7 ml of blue 5x TBE gel mix. Two air bubbles were passed through the pipette to mix the contents slightly, and the mix then carefully pipetted into the gel sandwich followed immediately by the remaining 34 ml 0.5x TBE gel mix. A plastic card gel comb was placed in the top of the gel and the gel mix left to polymerise. Up to a maximum of 32 (8 x 4) samples could be loaded on one gel.

2.2.7 Immobilisation of DNA on Nitrocellulose Membranes

2.2.7.1 Southern Transfer from Agarose Gels

Method of Southern (1975)

The gel was first stained with ethidium bromide and photographed to enable a later comparison with the post-blotted gel to give an indication of the extent of transfer. To facilitate the transfer of the larger DNA fragments the DNA was nicked prior to transfer. This was effected by partial depurination of the DNA. The gel was submerged in two changes of 0.25M HCl for 10 to 15 min (Wahl *et al.*, 1979), rinsed with water then soaked twice in 250 ml of denaturing solution (0.5M NaOH, 1M NaCl) for 30 min each. The gel was then soaked in neutralising solution (1.5M Tris-HCl,pH 6.5, 3M NaCl) for 60 min followed by a 10 min wash in 2x SSC buffer. The gel was then placed on a 'wick' of Whatman 3MM filter paper (soaked in 20x SSC) draped over a glass plate resting on top of a plastic tray containing 1 litre of 20x SSC transfer solution. The ends of the wick were immersed in this buffer. The edges of the gel were sealed from contact with the wick by sliding strips of X-ray film underneath. A sheet of nitrocellulose membrane, soaked in 2x SSC for 10 min, was placed on top of the gel, care being taken not to trap air bubbles between the two surfaces. Three sheets of Whatman 3MM filter paper, cut to the size of the gel and were soaked in 2x SSC, and placed on top of the gel, and then a wad (10 cm) of blotting paper placed on top. A small weight (approximately 250-500g) was placed on top to ensure an even contact between the gel and the membrane. Transfer was allowed to proceed for between 18 to 36 hours, after which time, the membrane was removed, washed briefly in 2x SSC, dried at ambient temperature and then baked at 80^{o}C for 2 hours under vacuum. The

gel was restained with ethidium bromide and photographed to visualise the extent of transfer of DNA.

2.2.7.2 *In Situ* Colony Hybridisation

The method employed was a modification of that described by Grunstein and Hogness (1975).

The bacterial colonies to be screened were replica-plated onto the ordered grid pattern of a 9 cm nitrocellulose filter discs (Schleicher & Schuell) placed on the surface of agar plates containing the selective antibiotic and in identical positions on agar master plates also supplemented with the selective antibiotic. Up to 100 colonies could be streaked on a single filter. Following overnight incubation at 37°C, the master plates were stored at 4°C and the filters peeled off the replicas. The cells were lysed using the following steps, these were performed by placing the filters, colony sides up, onto trays containing 3 layers of Whatman 3MM filter paper soaked in the appropriate solution. The filters were treated with 0.5M NaOH for 10 min, followed by 3 treatments with 1M Tris-HCl,pH 7.4 for 5 min each. Finally the filters were treated with 1.5M NaCl; 0.5M Tris-HCl,pH 7.4 for 10min. The filters were then immersed in 1x SSC, supplemented with 1 mg ml^{-1} proteinase K and 0.5% (w/v) SDS, and incubated at 37°C for 1 hour. The filters were then rinsed in 2x SSC, air dried, and baked in a vacuum at 80°C for a minimum period of 30 min.

2.2.8 Radio-Labelling of DNA

2.2.8.1 5′ -End Labelling of Synthetic Oligonucleotides

Synthetic oligonucleotides were supplied deprotected at the 5′-hydroxy terminus and therefore in a condition ready for phosphorylation at the 5′-hydroxy terminus with [γ-^{32}P] dATP by T4 polynucleotide kinase (Maxam & Gilbert, 1977).

Typically, 30 µl [γ-^{32}P] ATP (3000Ci $mmol^{-1}$; 50% (v/v) ethanol) was completely evaporated in a 1.5 ml Eppendorf tube under vacuum. To this, 150ng oligonucleotide probe, 3 µl of 10x kinase reaction buffer (2.1.4), and 1 µl 100mM DTT was added and

the reaction mixture made up to a total volume of 30 µl with distilled water. Following the addition of 0.2 units of T4 polynucleotide kinase, the mixture was incubated at 37°C for 30 min. The reaction was stopped by the addition of 1 µl 0.5M Na_2EDTA and kept on ice or frozen at -20°C prior to use. It was not usually considered necessary to remove excess label by chromatography through Sephadex G-50 as the oligonucleotide probe was usually in a 1.5 molar excess over ATP.

2.2.8.2 End-Labelling of Lambda/*Hin*dIII DNA Size Markers

Lambda DNA *Hin*dIII fragments were used as size markers on agarose gels and radio-labelled *Hin*dIII fragments were necessary to show up on autoradiographs of Southern blots.

Approximately 10 µg of lambda DNA *Hin*dIII fragments were diluted to a total volume of 90 µl in sterile distilled water. To this was added 10 µl of 10xTM buffer (2.1.4), 1µl (10 µCi) [α-^{32}P] ATP and 2 µl Klenow DNA polymerase I. The mixture was incubated at room temperature for 30min. 50 µl of Ficoll tracking dye was added prior to storage at -20°C. Aliquots of about 4 µl were used directly for electrophoresis.

2.2.9 DNA-DNA Hybridisations and Autoradiography

2.2.9.1 Annealing of Oligonucleotide

The membrane was prewetted in 2 x SSC and then placed in a sealable perspex box and a sufficient volume of pre-hybridization solution (214) to cover the filter added (50 to 100 ml). Prehybridization was carried out at 65°C for 15 to 30 min. The 5'-end-labelled oligonucleotide probe was then added to a final concentration of 1 ng ml^{-1} (with a specific activity approximating 10^7 cpm ml^{-1}). Hybridization was carried out at the calculated dissociation temperature for the oligonucleotide less 5°C (T_d-5°C). The T_d was defined to be the temperature at which 50% of the duplexes dissociate in 6 x SSC (Wallace *et al.*, 1979) and calculated as follows:

$$T_d\ (°C) = 4(G+C) + 2(A+T)$$

Hybridization was normally carried out for 2 hours or until sufficient time had elapsed to enable the probe to reach a calculated value of 1 to 3 x $Cot_{1/2}$. $C_ot_{1/2}$ is the calculated time required for half-renaturation of the probe (Maniatis *et al*, 1982) and was calculated as follows:

$1/X \times Y/5 \times Z/10$ = number of hours to achieve $C_ot_{1/2}$

where,

X = The weight of probe added (in ug).

Y = The complexity of the probe (length of probe in kilobases).

Z = The volume of the hybridization reaction.

Following the hybridization the membrane was washed with several changes of 6 X SSC, 0.1% (w/v) SDS, for 5 min each at the hybridization temperature (unless otherwise stated). The membrane was finally rinsed with 6 X SSC at room temperature and either wrapped in saran wrap or air dried prior to autoradiography (2-2-9-3).

2.2.9.2 Re-hybridisation of Membranes

If a previously used membrane was to be re-hybridised with a different probe, the bound probe was removed by several washes with boiling 0.1x SSC, 0.5% (w/v) SDS, for 5 min each. The membrane was finally rinsed in 2x SSC at room temperature and air dried. The membrane was then examined by autoradiography to ensure that all the probe had been removed before proceeding with the re-hybridisation.

2.2.9.3 Autoradiography

Autoradiography was carried out with Kodak X-Omat S, X-ray film which was positioned on top of the membrane or immobilised gel within a lightproof cassette, and left from 1 to 72 hour at room tempertaure before developing. When necessary the sensitivity of the autoradiography could be increased by the following methods. i. a calcium-tungstate-phosphor intensifying screen (Dupont Cornex Lightning-Plus) was used in the assembly in the cassette (Swanstrom and Shank, 1978) and the

autoradiography carried out at -70°C. ii. the film was sensitised by "preflashing", with a filtered photographic flashgun, prior to autoradiogrpahy (Laskey and Mills, 1977), and iii. films with lower threshold were employed such as Kodak XAR5 and Amersham's Betamax film.

2.2.10 DNA Sequencing

The method employed was the "chain-termination" procedure first described by Sanger *et al.* (1977) using the large fragment of *E. coli* DNA polymerase I. DNA fragments to be sequenced were first cloned into the replicative form (RF) of the M13 cloning vectors mp8, and mp9. These vectors contain multiple cloning sites in *lacZ'*, coding for the α-peptide of β-galactosidase. Insertion of foreign DNA into this polylinker sequence destroys active α-peptide synthesis such that recombinants are easily identified on media containing the chromogenic substrate for β-galactosidase, X-gal (BCIG) as "white" plaques. Single-stranded template DNA is prepared from the mature phage particles of the recombinants and is used directly in the sequencing reaction. An oligonucleotide primer, which is complementary to a sequence contained within the viral DNA ("sequence universal primer") or the inserted DNA, is hybridised 3′ to the template. Thus the sequence complementary to this insert DNA can be determined by sequential extension with the Klenow polymerase in the presence of both deoxyribonucleotides and the chain-terminating dideoxyribonucleotide analogues. Subsequent size fractionation of the reaction products by denaturing polyacrylamide gel electrophoresis enables the DNA sequence to be read.

2.2.10.1 Sub-Cloning of DNA into M13 Vectors

Routinely, 50ng of cleaved M13 vector was ligated with a two fold molar excess of target DNA fragments. The ligation products were used to transfect competent *E. coli* JM101 (2.2.2.). After transfection, the cells were added to a sterile universal, in a 47°C heating block, containing: 3 ml molten H-top soft agar , 20µl of IPTG (20 mg ml^{-1} in distilled water), 30µl of X-gal (40 mg ml^{-1} in DMF), and 200µl of exponentially

growing *E. coli* JM101 cells. The mixture was poured onto a 2xYT agar plate and allowed to set before the plate was inverted and incubated at 37°C overnight. Recombinants were identified as colourless plaques owing to the insertional inactivation of the functional β-galactosidase α-peptide encoding sequence.

2.2.10.2 Sonication of DNA

Sonication of DNA is one of the methods used to generate the random sized fragments of DNA used in the "shotgun" approach to sequencing large DNA molecules. These random fragments were then cloned into M13 vectors (Messing *et al.*, 1977; Messing & Vieira, 1982). These are then sequenced and the data compiled to ultimately reveal the complete sequence. The sonication method followed was that of Deininger (1983), using an MSE soniprep 150 sonicator.

The sonication probe was sterilised by boiling in 1% (w/v) SDS for 15 min, rinsed well in ethanol and chilled to -20°C in ethanol. 10μg of the purified DNA fragment made up to a volume of 100 μl with 1x TE buffer. This was kept on ice throughout the procedure. The probe was inserted into the Eppendorf tube as far as possible without touching the tube and allowed to cool for 10 min by contact with the ice. With the sonicator switched to maximum output, and an amplitude setting of 3, four 8 sec bursts, with 1 min cooling periods between each, were applied to the DNA solution. The solution was recovered and adjusted to a volume of 150μl with 20μl 10x T4 DNA polymerase buffer (2.1.4), 20μl dNTP mix (dATP, TTP, dCTP, and dGTP; 0.25mM each), 5μl T4 DNA polymerase, and 1μl Klenow polymerase. The reaction mixture was then incubated at 14°C for 3 to 4 hours. The sample was electrophoresed on a 1.5% (w/v) agarose mini gel (2.2.6.1), alongside DNA size markers, until the Bromophenol Blue dye front had migrated 2 to 3 cm from the well. DNA fragments between 400 and 800 bp in length were excised from the gel and the DNA extracted by electro-elution (2.2.4.6), and ethanol precipitated (2.2.4.7.3).

2.2.10.3. Preparation of Template DNA

Recombinant plaques (colourless), were picked into 2 ml aliquots of 2xYT broth, containing 100 µl per 100 ml of an overnight culture of *E. coli* JM101, and incubated at 37°C for 6 hours with vigorous aeration. The cultures were then transferred to Eppendorf tubes and centrifuged (Sorval RC5B, SHTM rotor, 13000 r.p.m. for 10 min at 4°C). The supernatants (1.25 ml vol) were decanted into fresh Eppendorf tubes containing 250µl 20% PEG (6000), 2.5M NaCl, and the contents mixed well by inversion. After being left to stand for a minimum of 15 min, the tubes were again centrifuged (Sorval RC5B, SHTM rotor, 13000 r.p.m. for 10 min). The supernatants were discarded by aspiration leaving the phage pellets in the tubes, these were briefly centrifuged for 30 sec as above, and all traces of supernatant discarded by careful aspiration. The phage pellets were resuspended in 150 µl volumes of 1xTE buffer (using an Eppendorf shaker) and extracted once with a 50 µl of TE saturated phenol. Following separation of the emulsions by centrifugation (MSE Microcentaur, 13000 r.p.m. for 1 min at room temperature), 130 µl of the aqueous phase (top phase) was transferred to fresh Eppendorf tubes making sure not to transfer any of the bottom phenol layer, any traces of phenol were removed by extracting 4 times with 0.5 ml of diethyl ether. The samples were then ethanol precipitated (2.2.4.7.3.), washed with 70% (v/v) ethanol, dried under vacuum, and resuspended in 30µl 1x TE buffer.

2.2.10.4 Methods used for the dideoxy-sequencing involving α-^{32}P-dATP (used for sequencing the *S. aureus spa* gene) Modification by Coulson and Winter, MRC Cambridge (personal communication)

2.2.10.4.1 Annealing of Template and Primer

The templates to be sequenced were first annealed with the appropriate oligonucleotide primer. This was carried out in a drawn out glass capillary tube. 5 µl of template DNA, 1 µl of 10 x annealing buffer, 1 µl of primer (2 ng μl^{-1} in double distilled Millipore water) and 3 µl of double distilled water were drawn up into a capillary tube

and mixed by pipetting in and out of the capillary several times. The capillary tube was then heat sealed at both ends and the contents annealed by incubation in a test tube of boiling water placed in a boiling water bath for 3 minutes, after which the test tube was removed from the water bath and allowed to cool slowly to room temperature.

2.2.10.4.2 Polymerisation (sequencing reactions)

These reactions were performed in autoclaved capless 1-5 ml Eppendorf tubes, the tubes were positioned in ordered racks (in an equivalent to a 4 x 4 matrix) volumes stated are sufficient for 4 templates.

Four 5 µl aliquots of a α-32P-deoxyadenosine triphosphate (specific activity 400 ci/mmole were dried down using a vacuum line. 10 µl of C termination mix, (termination mix is made up of 5 µl of the corresponding ddNTP and $dNTP^{O}$ working solutions, (section 2.1.4.1)) was added to the first tube of dry 32P-dATP and mixed gently to dissolve, this procedure was repeated with the T, A and G termination mixes. The capillary tubes containing the annealed template and primer were cut open and 1 unit of Klenow polymerase drawn into each, this was mixed and then blown into the bottom of an Eppendorf tube and the capillary discarded. For each template, 2 µl of the DNA/polymerase mix was transferred to the inside wall of 4 individual tubes (forming the horizontal row) these being for the 4 independent sequencing reactions. Then 2 µl aliquots of the T,C,G and A termination 32P-dATP mixes were added to the inside wall of the appropriate reaction tubes (forming the vertical rows) care was taken not to mix this with the aliquoted template. The reactions were started by mixing using a brief centrifugation (5 sec, IEC Centra 4X with centra rack rotor) and the reaction allowed to proceed for 15 min at room temperature. During the reaction, a 1 µl aliquot of 0.5 mM dATP chase was aliquoted to the inside wall of the reaction tubes, this was mixed by brief centrifugation. Following a further 30 min incubation at room temperature, 4 µl of formamide dye mix was added to each tube and mixed by centrifugation. If the samples were not to be run on a gel immediately they were frozen without the addition of the formamide dye mix.

2.2.10.5 Method use for dideoxy-sequencing involving α-^{35}S-dATP modification communicated by Bankier (1984) MRC Cambridge (personal communication)

2.2.10.5.1 Annealing of template and primer

The single stranded template and appropriate oligonucleotide primer were first annealed. This was carried out in capped Eppendorf tubes. 5 µl of templates DNA was mixed with 1 µl TM buffer, 3.5 µl distilled water, and 0.5 µl primer (2 ng μl^{-1} in oligonucleotide storage buffer), and annealed by incubation of the tube in an 80°C water bath for 3 min, after which it was allowed to slowly cool to room temperature. The tube was then briefly centrifuged (IEC Centra 4X with centra rack rotor) for 5 sec to ensure recovery of the contents in the bottom of the tube.

2.2.10.5.2 Sequencing reactions (polymerisation)

These reactions were performed in autoclaved capless, 1.5 ml Eppendorf tubes positioned in ordered centra racks.

For each template, 2 µl of the annealed mix was transferred to the inside wall of 4 inidividual tubes; these being for the four independent sequencing reactions: with the "T", "C", "G", and "A" deoxy/dideoxynucleotide mixes which also contained [α-^{35}S] dATP (0.125u Ci μl^{-1}, specific activity 400 ci $mmol^{-1}$) and Klenow polymerase (0.125 u μl^{-1}). A volume of 2 µl aliquots of these nucleotide/polymerase mixes were transferred, onto the inside wall aliquoted template. The reactions were started by brief centrifugation (IEC Centra 4X with centra rack rotor) and the reaction allowed to proceed for 15 min at room temperature. 2 µl aliquots of sequence "chase" (0.25 mM dNTPs were then added to the inside wall of the reaction tubes and again mixed by centrifugation. Following a further 15 min incubation at room temperature, 2 x 2 µl of formamide dye was added to each tube and mixed by centrifugation.

2.2.10.6 Loading and running of Sequencing Gels

Denaturing polyacrylamide gradient gels (2.2.6.2) were routinely used. The order of loading of the reaction mixes for each template was T,C,G,A. Immediately before loading, the samples were heated in a boiling water bath for 2 min and the gel wells flushed out with 1x TBE running buffer. A 2.5 µl sample was loaded in each well using a drawn out glass capillary and a constant current of 25 mA was applied. The gels were normally run for 3.5 hours, by which time the Bromophenol Blue tracking dye had just run off the bottom of the gel. After electrophoresis, the gel, adhering to the back plate was fixed by immersion in 10% (v/v) glacial acetic acid, 10% (v/v) methanol, for 15 min. After being left to stand in a fume cupboard for 1 hour the gels were dried by incubation at 80°C for 2 hours, and then autoradiographed (2.2.9.3)

2.2.10.7 Compilation and Analysis of Sequencing Data

All computer analyses performed on sequence generated for the SpA DNA were undertaken using the programmes developed by Staden (1980). All computer analyses performed on sequence generated for the alk phos DNA were undertaken using the programmes supplied by DNASTAR Inc (Madison,U.S.A.).

2.2.11 **Plasmid Sequencing**

Preparation of Single Stranded Template from Plasmid DNA

Distilled water was added to 4 µg of plasmid DNA to give a total volume of 16 µl. The double stranded plasmid was denatured with 4 µl of a freshly prepared solution of 1 M NaOH/1 mM EDTA and incubated at room temperature. After 5 min, the microeppendorf was placed on dry ice, and neutralised with pre-chilled 2 µl of 2 M ammonium acetate, pH 4.5. Immediately 60 µl of pre-chilled 10% ethanol was added, and the microeppendorf placed at -20°C. After 1 hour, the microeppendorf was spun for 20 min in an Eppendorf microfuge (MSE Microcentaur, 13000 r.p.m.), the pellet washed with 60 µl pre-chilled 80% (v/v) ethanol, and the pellet briefly dessicated for 2 min. The dessicated pellet was dissolved in 7 µl double distilled water, and used

directly for primer annealing (2.2.10.5.1).

2.2.12. Site-Directed Mutagenesis

The method employed was a derivation of that described by Carter *et al.* (1985).

The DNA to be altered was first cloned into a suitable M13mp vector. The recombinant M13 template DNA was then annealed with an appropriately phosphorylated mutagenic oligonucleotide. A subsequent Klenow polymerase-directed 'extension reaction', primed by the phosphorylated mutagenic oligonucleotide, in a repair-deficient *E. coli* host, resulted in the generation of a mutant (-) M13 strand. This (-) strand was then copied by the compromised host to give the mutant (+) M13 strand. Resultant M13 clones were screened by hybridisation experiments with the mutagenic oligonucleotide, and those mutants identified, subjected to DNA sequence analysis.

2.2.12.1 Phosphorylation of Mutant Oligonucleotides

The freshly synthesised mutagenic oligonucleotide, which had been deprotected, was adjusted to a concentration of 100ng μl^{-1}. A 10µl aliquot of the oligonucleotide was mixed with 2µl 10x kinase buffer (2.1.4), 1µl 100mM DTT, 2µl 10mM rATP, 1µl T4 polynucleotide kinase, and 4µl distilled water. The reaction mixture was incubated at 37°C for 30 min and then the remaining enzyme inactivated by heat treatment (70°C for 10 min).

2.2.12.2 Extension Reactions and Transformation of *E. coli* BMH71-18

The M13 template DNA (DNA to be altered) was adjusted to an approximate concentration of 2 µg in 15 µl (approximately 20 pmol) in 1xTE buffer. 15 µl of this DNA was added to a tube containing 4 µl of the phosphorylated oligonucleotide (2.2.8.1) and 2 µl of 10xTM buffer, and annealed as described in 2.2.9.1. Once the annealed template had cooled to room temperature the following were added: a further 2 µl 10xTM buffer (2.1.4), 4 µl of a dNTP solution (2.5 mM dATP, TTP, dCTP, and dGTP), 1 µl 10 mM rATP, 2µl 100mM DTT and 10µl distilled water. The reaction

mixture was placed on ice and then 2 µl T4 DNA ligase (2.5 units μl^{-1}) and 0.8µl Klenow polymerase (BCL, 5 units μl^{-1}) was added and the mixture incubated at 15°C for 16 hours. The reaction was stopped with the addition of 200µl 10mM Na_2EDTA, 100 mM Tris-HCl,pH 8.1.

A range of aliquots (1, 5, and 10µl) of the arrested extension reaction mixture was used to transform 200 µl of amounts of competent *E. coli* BMH 71-18 *mut*L (2.2.2). The transformed cells diluted 10 and 100 times and added to 3 mls of molten (45°C) H-Top agar overlays, which were supplemented with 200 µl of exponentially growing *E. coli* JM101, were poured over 2xYT agar plates, and allowed to set. The plates were incubated overnight at 37°C.

2.2.12.3 Colony Screening of Mutants

For each mutagenesis experiment 100 M13 plaques were picked and replica picked onto 2xYT agar plates in an identical asymmetric pattern, allowing the correct alignment of positive mutants to colonies on the master plate. The plates were incubated overnight at 37°C and then the plaques removed as follows a 9 cm nitrocellulose disc was placed gently on the surface of the agar plate, on which the phage-infected bacterial colonies had grown, and left to stand at room temperature for 1 min. The filter was carefully lifted away, with the bacterial colonies adhering to it, and processed as described for *in situ* colony hybridisation (2.2.7.2) with the omission of the 2x SSC rinse and proteinase K treatment. The filter was placed on a sheet of Whatman 3MM filter paper, colony side up, and allowed to air dry at room temperature. The filter was then placed between two glass sheets to prevent it curling up and baked at 80°C in a vacuum oven for 30 min.

The filter was prewetted in 6x SSC for 5 min and hybridised with the mutagenic oligonucleotide which had been end-labelled with [γ-^{32}P] dATP (2.2.8.1). The screening was performed exactly as described in section (2.2.9) with the exceptions that the sheared heat denatured salmon sperm DNA was omitted from the pre-hybridisation

solution and that the hybridisation was carried out at a T_d-15°C for the oligonucleotide. The filter was washed with 3 x 1 min changes of 100 ml 6x SSC at room temperature, covered in saran wrap and autoradiographed (2.2.9.3) at - 70°C for 1 hour. This revealed the asymmetric pattern of the bacterial colonies. The filter was then progressively washed (with 100 ml 6x SSC) at increasing temperatures, with subsequent autoradiography of the filter in between each wash, until the T_d (°C) of the mutagenic oligonucleotide had been reached. As the T_d temperature was approached the hybridisation signal of the non-mutant colonies decreased in strength relative to that of the positive mutant colonies (as a consequence of the lower T_d due to the mismatch). At the T_d of the mutagenic oligonucleotide (or in some cases as a few degrees higher) the only hybridisation signals obtained were from the positive colonies.

2.2.12.4. Plaque Purification and DNA Sequencing

Individual mutants were obtained by the following procedures; single positive colonies were transferred to 10 ml of 2xYT-broth and vortexed. A 10 µl volume of these diluents was added to 3 ml molten (47°C) H-top agar containing 200 µl exponentially growing *E. coli* JM101 and poured onto a 2xYT agar plate to set, and incubated overnight at 37°C. Isolated plaques were chosen and template DNA prepared from them (2.2.10.3). The entire cloned insert of each template was subjected to DNA sequence analysis as described in 2.2.10.6. This was generally accomplished by using a number of custom synthesised oligonucleotide primers. DNA sequence analysis not only confirmed the mutated sequence but also checked that no other sequence alterations had occurred. Template DNA (1µl) of the chosen mutant(s) was used to transfect competent *E. coli* JM101 cells (2.2.2) and the transfected cells used to inoculate 10 ml 2xYT-broth for overnight incubation to provide the 'phage inoculum' for the isolation of double-stranded RF DNA.only confirmed the mutated sequence but also checked that no other sequence alterations had occurred. Template DNA (1µl) of the chosen mutant(s) was used to transfect competent *E. coli* JM101 cells (2.2.2) and the transfected cells used to inoculate 10 ml 2xYT-broth for overnight incubation to provide the 'phage inoculum' for the isolation of double-stranded RF DNA.

2.2.13 Protein Analysis

2.2.13.1 Protein Assay

Protein concentrations were determined by the Folin method of Lowry *et al.* (1951), with bovine serum albumin as the standard.

2.2.13.2 Preparation of Cell Free Lysates from *E. coli* and *S. aureus* for Detection of SpA by Western Blotting or Rocket Immunoelectrophoresis

A single colony was inoculated into 10 ml of L-broth containing an appropriate antibiotic and grown at 37°C overnight. The cells were harvested by centrifugation (Sorval RC5B SS24, rotor), 6000 rpm for 10 minutes at 4°C, the supernatant removed and reserved (for later analysis for SpA) and the cells resuspended in 1.0 ml of 50 mM Tris HCl, pH 8.0, 25% sucrose. For *E. coli* cells 0.15 ml of lysozyme (20 mg ml^{-1} in 250 mM EDTA pH 8.0) was added and the cells left on ice for 20 minutes, swirling occasionally. For *S. aureus* cells 0.15 ml of lysostaphin (200 mg ml^{-1} in 250 mM EDTA pH 8.0) was added and the cells left on ice for 5 minutes, swirling occasionally. Then 1.0 ml of 250 mM EDTA, pH 8.0, was added and the cells replaced on ice, swirling occasionally. After 5 mins 1.5 ml of Brij/Doc solution (2.1.4) was added, mixed thoroughly, the left on ice for a further 20 to 30 minutes or until lysis had occurred. The cell debris and bulk of DNA was then pelleted by centrifugation (Sorvall RC5B, SS24 rotor, 15000 rpm for 45 mins at 4°C) and the cleared lysate recovered.

2.2.13.3 Preparation of cell free sonicates (*E. coli*) for detection of SpA by western blotting and immunoelectrophoresis

The cells were grown and harvested as for section 2.2.13.2. They were then resuspended in 3 ml of 50 mM Tris HCl pH 8.0, 25% sucrose. The cell suspension was transferred to an MSE sonication tube and sonicated using a MSE soniprep 150 sonicator, maximum output, amplitude 2, for two sec bursts, with a 1 minute cooling interval at 4°C between bursts. The cell debris was then pelleted by centrifugation

(Sorvall RC5B, SS24 rotor, 15000 rpm for 45 mins at 4°C) and the cleared sonicate removed.

2.2.14 Electrophoresis of Proteins

2.2.14.1. SDS-Polyacrylamide Gel Electrophoresis (PAGE)

Proteins were separated by electrophoresis using polyacrylamide gels under denaturing conditions. The proteins were denatured by heat in the presence of the anionic detergent SDS and the reducing agent β-mercaptoethanol before they were loaded onto the gel. The denatured polypeptides bind SDS and become negatively charged. Because the amount of SDS bound is almost always proportional to the relative molecular mass of the polypeptide (1g protein to 1.4g SDS) and is independent of its sequence, SDS-polypeptide complexes migrate through polyacrylamide gels in accordance with the size of the polypeptide.

SDS-polyacrylamide gel electrophoresis was carried out with a discontinuous buffer system in which the buffer in the reservoir was of a different pH and ionic strength from that used in the gel. For greater resolution, a stacking gel of high porosity was cast on top of the resolving gel.

SDS-polyacrylamide gel electrophoresis was performed with 12.5% (w/v) slab gels (linear range of separation approximately 14-60 kDa) run in an LKB vertical Electrophoresis Unit (Laemmli,1970). The gel was poured in two stages, the stacking gel being cast after the resolving gel had polymerised.

The resolving gel was prepared with 14.5 ml of 30% (w/v) acrylamide stock, 17.25 ml of resolving buffer and 5.6 ml of distilled water. The mixture was deaerated by partial vacuum and polymerisation initiated with 15 µl TEMED and 0.4 ml of a 10% (w/v) ammonium persulphate solution. Two glass plates were cleaned with detergent and polished with absolute alcohol, greased spacers (silicone grease) were placed on the sides of the "front" (indented) plate and the top plate placed on this and clamped into

the running apparatus. This was then sealed with 2% hot agarose in SDS running buffer and allowed to stand for 30 min. The acrylamide solution was poured between the two glass plates leaving sufficient space for the stacking gel (the length of the teeth of the comb plus 1cm) to be poured. The gel surface was overlayed with isobutanol, and the gel left at room temperature to polymerise. The overlay was poured off and the gel surface washed with distilled water. The stacking gel was prepared with 3.0 ml of 30% (w/v) acrylamide solution, 12 ml of stacking buffer and 8 ml of distilled water. The mixture was deaerated by suction and polymerisation initiated with 20 µl TEMED and 200 µl of a 10% (w/v) ammonium persulphate solution. The stacking gel solution was poured directly onto the surface of the polymerised resolving gel. A teflon comb was then immediately inserted into the gel, being careful not to trap any air bubbles. Following polymerisation, the teflon comb was removed, and the wells washed out with distilled water. The wells were filled up with electrophoresis buffer and the samples loaded using a Hamilton microliter syringe (the samples were previously solubilised in at least 25% (v/v) of SDS loading buffer and heated at 100°C for 3 min). The gel was then mounted in the electrophoresis apparatus. The electrophoresis buffer was added to the top and bottom reservoirs, removing any air bubbles that became trapped at the bottom of the gel between the glass plates. Electrophoresis was carried out with a current of 12mA overnight, or until the Bromophenol Blue tracer had migrated off the end of the gel. Following electrophoresis, the gels were removed from between the plates and used for western blotting (2.2.15).

2.2.14.2 Rocket Immunoelectrophoresis

Electroimmunoassay introduced by Laurell, (1965), combines the speed of electrophoresis with the quantification of antigen provided by single radial immunodiffusion. In this modified version, the IgG is incorporated in the gel, and a set of protein A standards and cell lysates or supernatants containing SpA placed in wells cut into the gel. A direct current is applied and the SpA migrates into the gel according

to its charge. While the SpA is in excess it will form soluble complexes in the gel, however at equivalence it will form a precipitate in the gel, as additional SpA arrives, the precipitate at the advancing front dissolves, while the precipitate along the sides remains. When no more SpA is left to enter the precipitate at the advancing front, the lines of precipitate converge producing a rocket-shaped appearance (monorockets). The area enclosed by the precipitation lines or, more roughly the height of the rocket is directly proportional to the initial concentration of SpA in the well.

Monorocket immunoelectrophoresis was performed with a 1% agarose/Tricine buffer slab gel placed on an LKB cooling plate, within a modified horizontal electrophoresis tank, connected to a Grant Thermocirculator containing antifreeze.

The immunogel was prepared with a 15 ml volume of 1% agarose in tricine buffer, this was allowed to cool to 45°C and 7.5 µl of a 4.5 mg m^{-1} (in water) stock of human IgG (Sigma) added to give a final concentration of 22.5 µg ml^{-1}. The IgG containing agarose was then poured into a 10 x 10 cm perpex gel template, (which had previously been placed onto a levelled cooling plate and its edges sealed with 1% agarose), spread to the edges to give a final depth of approximately 2 mm and left to set. The required number of wells, (no less than 0.75 cm apart) were punched into the gel with a 3 mm diameter punch, about 1.5 cm from the cathode end of the gel. The template was removed and the cooling plate and gel transferred to the electrophoresis tank. 1x tricine buffer was added to the electrophoresis tank and used to soak 2 wicks cut from Whatman 3FF, these wicks were placed in contact with edge of the gel and the buffer tanks. The thermocirculator was connected to the cooling plate and set at 5°C.

A low electric current was maintained during sample loading to prevent diffusion and immunoprecipitation, and thus loss of sample, in the "wrong" direction. A known volume (usually 6 µl) of standards (made from a stock protein A solution, 0.64 mg ml^{-1} in water, diluted with tricine buffer to give 8 standard concentrations ranging from 20 µg ml^{-1} to 204 µg ml^{-1}) and samples were loaded into the wells. The voltage was then

increased to obtain a potential difference of 200 V across the gel, and electrophoresis carried out for 6 hours. Following the electrophoresis, the gel was removed from the cooling plate and soaked in a 0.85% sodium chloride solution for 30 minutes to remove any unprecipitated protein from the gel. The gel was then washed 2 x 1 min in running water and then stained by soaking in staining solution (2.1.4) for 5 minutes prior to destaining in clearing solution (2.1.4). The monorockets were measured and a calibration curve produced by plotting distance travelled by the rocket against the known standard SpA concentrations. From this the unknown SpA concentration in the cell samples could be determined.

2.2.15 Immobilisation of Protein on Nitrocellulose (Western Blocking)

Proteins were applied to nitrocellulose membranes by electrophoretic transfer from SDS-polyacrylamide gels as described by Towbin *et al.*(1979).

One side of the gel was placed in contact with a nitrocellulose filter. The gel and its attached filter were then sandwiched between 3 sheets of Whatman 3MM paper, two porous pads, and two plastic supports. The entire construction was then immersed in the electrophoresis tank containing the electrophoresis buffer. The nitrocellulose filter was placed towards the anode and a voltage of 60V applied for 4 hours, during which time the proteins became attached to the filter. Once transfer had occurred, the nitrocellulose filter was probed (2.2.17).

2.2.16 Iodination of Proteins

Proteins electrophoretically transferred to nitrocellulose filters were probed with iodinated proteins. Proteins were iodinated using a reaction in which $H_2O^{125}I^+$, generated by oxidation of $Na^{125}I$ with chloramine T, attacks the side chains of tyrosine residues and,to a lesser extent, histidine residues (Greenwood *et al.*, 1963). This modification, if not carried out to excess, generally does not affect the protein to protein interaction. For safety reasons, iodination was performed in a fume cupboard.

100 µg of human IgG was dissolved in 100 µl pBS in a 1.5 ml sterile Eppendorf tube. The iodinating conditions were created by the addition of 20µl 10mM chloramine T (freshly prepared in 1xPBS buffer) followed by 0.2 µl of $Na^{125}I$ (~200µCi). Iodination was allowed to proceed at room temperature for 3 min. Iodination was stopped by the sequential addition of 20 µl 10mM potassium metabisulphite, 60 µl 10mM potassium iodide and 100 µl of 10% w/v BSA (RIA trade) in 1xPBS buffer.

The iodinated protein was separated from unincorporated iodine using a G-50 column (packed bed volume of approximately 25 ml). The column was equilibrated with 10% BSA, 0.1% Triton X-100 in 25 ml of PBS (running buffer). The iodinated probe was added to the column and eluted with running buffer. The first activity peak (total vol 1.0 to 1.5 ml) was collected and used in probing the western blotted filters.

2.2.17 Probing of Western Blot Filters

Once transfer had occurred, the nitrocellulose filter was probed by the following method. The filter was soaked in 100 ml of 5% BSA (RIA grade), 1.0% Triton x 100 in phosphate buffered saline (PBS) for either 2 hours on a rotating platform at room temperature or overnight at 4°C, to saturate non-specific protein binding sites. The human IgG/I^{125} labelled probe was added and incubation continued either at room temperature for 4-6 hours or overnight at 4°C. The hybridising buffer was then carefully removed and the filter rinsed in 6 changes of buffer; 2 x 30 mins in 5% BSA, 1% Triton x -100 in PBS, then 2 x 30 mins in 1% Triton PBS; and finally 2 x 30 mins in PBS. The filter was removed and dried on Whatman 3 mm paper.

2.2.18 Enzyme Localisation in *E. coli*

The localisation of SpA within the *E. coli* cell to a particular 'cell-fraction': (cytoplasm, periplasm or membrane) was determined using an adaptation of the traditional cell fractionation technique of Heppel, 1967.

2.2.18.1 Preparation of cell fractions

A 50 ml culture of the *E. coli* clone under investigation was grown in L-broth (Ap 50 $\mu g\ ml^{-1}$) overnight at 37°C in an orbital shaker. A 2 ml aliquot was inoculated into 100 mls of low phosphate broth (LPB) and grown as above until an OD_{450} of 1.0 was reached. The culture was then split into 3 amounts, a: 12 ml, b: 24 ml and C: 40 ml; and the cells harvested from each by centrifugation (8000 rpm in a Sorvall RCBS, SS34 rotor, 10 mins at 4°C). The supernatant was discarded and the cells washed by resuspending in a 5 ml of 10 mM Tris pH 7.0 followed by centrifugation (as above). Each of the 3 different harvest volumes were then treated separately.

The 12 ml cell culture volume was resuspended in 1.5 ml of 10 mM Tris pH 7.0 0.2 mM DTT and stored on ice until assayed, this gave an indication of the amount of cell lysis.

The 24 ml cell culture volume was resuspended in 3 ml of 10 mM Tris pH 7.0, 0.2 mM DTT and the cells sonicated using an MSC soniprep 150 sonicator (3 x 20 sec bursts, set on medium frequency output, amplitude 2, with a 45 second cooling interval at 4°C between bursts). The cells were stored on ice until assayed, this would indicate the total cell enzyme content.

The 40 ml cell culture volume underwent cell fractionation in the following manner. The cell wall was disrupted by a mixture of osmotic shock and gentle lysis using lysozyme by resuspending the cells in 0.9 ml of 20% sucrose, 30 mM Tris pH 8.0, 0.2 mM DTT, to which 20 µl of lysozme (2 mg ml^{-1} dH_2O) was added followed immediately by 40 µl of 0.1M EDTA pH 8.0. This was then incubated for 10 mins at 23°C and placed on ice where 300 µl of 10% BSA was added, followed by the addition of 5 ml of sucrose-tris-DTT (as above) this was mixed by 4 gentle inversions of the tube. The periplasmic fraction of the cell was then separated by centrifugation (8000 rpm in a Sorvall RCBS, using a SS34 rotor, 10 mins at 4°C) the supernatant

(periplasmic fraction) removed and placed on ice. The cell pellet was drained and resuspended in 5 ml of 10 mM Tris pH 7.0, 2 mM DTT and the cells disrupted by sonication (MSE soniprep 150, 15 seconds at low frequency, amplitude 2). The sonicate was then centrifuged at 3000 g for 10 min at 4°C (the pellet containing cell debris discarded) and the supernatant containing the cytoplasmic and membrane fractions placed in a 13.5 ml ultracentrifuge tube. The volume was made up with 10 mM Tris pH 7.0, 0.2 mM DTT and then centrifuged in a Beckman OTD 75 ultracentrifuge, Sorval Ti50 fixed angle rotor at (50,000 rpm) 4°C for 50 mins. The supernatant (cytoplasmic fraction) was carefully poured off and placed on ice until assayed. The pellet (membrane fraction) was drained and resuspended in 1 ml of 10 mM Tris pH 7.0, 0.2 mM DTT.

2.2.18.2 Detection of marker enzymes

Change in absorbance was measured using a Beckman spectrophotometer U8.

2.2.18.2.1 Alkaline Phosphatase (periplasmic)

To each 100 µl aliquot of extract, 900 µl of nitrophenyl phosphate (0.8 mg ml^{-1} in 1M Tris, pH 8.0) was added and the change in absorbance Δ E at 420 nm measured as the mixture incubated at 37°C. If the reaction was too rapid the cell extract was diluted in 1M Tris pH 8.0 and the assay repeated. Units of activity per ml were determined using the following calculation:

$$\text{units/ml} = \frac{\text{final sample volume (1 ml)} \times \Delta E \text{ min}}{E\,(18.5) \times \text{path length (1cm)} \times \text{actual sample vol (sample x dilution factor)}}$$

2.2.18.2.2 NADH oxidase

To each 100 µl aliquot of cell extract, 900 µl of assay buffer 0.7 ml Tris-HCl (50 mM pH 7.5) 0.1 ml NADH (0.12 mM), 0.1 ml DTT (2 mM) 0.1 ml was added and the change in absorbance Δ E at 340 nm measured as the mixture incubated at 37°C. If the reaction was too rapid the cell extract was diluted in 0.1 M Tris pH 7.5. Units of activity m^{-1} were determined using the same formula as for alkaline phosphatase activity except the value of E = 6.22.

2.2.18.2.3 Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)

To each 100 µl of extract 1.5 ml of assay mix was added (section 2.1.4) and the change in absorbance Δ E at 340 nm measured as the mixture incubated at 37°C. If the reaction was too rapid the cell extract was decreased (100 µl - 5 µl). The units of activity ml^{-1} were determined using the same formula as quoted for NADH-oxidase, except the total sample volume variable.

2.2.18.2.4 SpA

The amount of SpA present in each sample was determined using neat or diluted cell extract and monorocket electrophoresis (2.2.14.2)

2.2.19 **Synthesis of Oligonucleotides**

Oligodeoxyribonucleotides (oligonucleotides) were synthesised by 'solid phase oligonucleotide synthesis' using an Applied Biosystems Model 380A DNA synthesiser. The coupling chemistry employed was the phosphoramadite method which has inherently high coupling yields, typically 95-100%. The principles of this method of synthesis have been described elsewhere (Gait *et al.* ,1982).

2.2.20 **N-Terminal Residue Analysis of Protein**

Automated Edman degradation was introduced in 1967 (Edman and Begg,1967). Reagents in the gas-phase were introduced to replace liquid-phase reagents

(Hunkapiller *et al.*,1983). The sequence of amino acid residues at the amino-terminal region of the protein was determined using automated Edman phenylthiohydantoin degradation with an Applied Biosystems Model 477A Pulse Liquid Protein Sequencer as described by Hunkapiller *et al.*,1983.

2.2.21 Light Microscopy

Samples were taken from exponentially growing cultures of *E. coli* JM83 containing each of the three pPA plasmids. Gram strains (Jensens modification 1984) were performed on each of the host/vector systems using *E. coli* JM83 as a negative control. Photographs were taken using a Leitz Dialux 20 microscope, with an automatic wild MPS51 film holder.

2.2.22 Electron Microscopy

Immuno-electron microscopy was used to analysis the effect of SpA on host cell structure and to determine the location of SpA within the cell.

2.2.22.1 Preparation of sections of host organisms for structural analysis

A sample from the continuous culture experiments was centrifuged (6000 rpm in a Sorvall RCBS SS34 rotor, at 4°C, 10 min) and fixed by resuspension in 2.5% glutaraldehyde in 0.1 M phosphate buffer for 2 hours at 4°C. This was followed by a buffer wash before post-fixation in 1% osmium tetraoxide in phosphate buffer for 2 hours at 4°C. The organisms were then pelleted (6000 rpm 10 min), the supernatant discarded and the pellet gently mixed with 2% molten agar (Oxoid No.1). When set, the agar was cut into approximately 1 mm cubes which were dehydrated in graded ethanol series (30%, 50% for 10 min. each: 70%, 90% for 30 min each) through to absolute ethanol (2 changes of 30 min) followed by 20 min (2 x 10 min changes) in 100% propylene oxide.

Embedding was carried out by infiltrating the agar cubes with 50:50 propylene oxide : araldite CY212 resin mixture for 4 hr followed by 24 hr in 100% araldite mixture. All of the above procedures were carried out at room temperature. The agar cubes were then placed in BEEM embedding capsules (size 00), covered with fresh araldite mixture and allowed to stand for a further 24 hour at room temperature before being cured at 60°C for 72 hour.

Thin sections (80 nm) were cut from the resulting blocks on a Reichert OMU2 ultramicrotone using a diamond knife and were picked up on 400 mesh copper EM specimen grids. After staining in aqueous uranyl acetate (2% for 30 min) followed by Reynolds lead citrate (10 min) (Reynolds, 1963), the sections were stabilised by carbon coating in an Edwards 12E6 vacuum coating unit prior to examination in a Phillips EM400T electron microscope operated at 80 Kv.

2.2.22.2 Preparation of sections of host organisms for the detection of SpA

Organisms were pelleted form the continuous culture medium by centrifugation (6000 rpm 10 mins at 4°C) and fixed by resuspension in 0.1% glutaraldehyde, in 0.1 M phosphate buffer for two hours at 4°C. The procedures followed in section A were then carried out, omitting the post-fixation of osmium tetraoxide.

2.2.22.3 Preparation of whole cells for the detection of SpA

A sample from the continuous culture medium was centrifuged (6000 rpm, 10 mins at 4°C) then resuspended and fixed in 0.1% formalin in 0.1 M phosphate buffer. The cells were then adhered to formvar/carbon filmed E.M. grids by leaving a drop on the surface for four minutes. The cells could then be probed for SpA as in 2.2.22.4.

2.2.22.4 Assay system for the detection of SpA using electron microscopy

The grids were first resuspended on a 150 µl droplet of new born calf serum for 30 min so as to inhibit any non-specific binding. They were then washed 3 x 5 mins in PBS (pH 7.6) and incubated for 1 hour on a 150 µl droplet of 1:500 (in PBS pH 7.6) of

rabbit anti-SpA for sections and 1:5000 for whole cells. The grids were then washed 3 x 5 min in PBS and blocked with new born calf serum for 30 mins. The grids were again washed in PBS (pH 8.2) then placed on 150 µl droplets of 1:10 (in PBS pH 8.2) goat anti-rabbit labelled with gold (Biocell Ltd.) for 1 hour. Finally the grids were washed 3x in PBS (pH 8.2) and dried on filter paper. All the above incubations were carried out at room temperature.

CHAPTER 3

3.1 INTRODUCTION

The *S. aureus* Cowan I *spa* gene was previously cloned by Duggleby and Jones (1983). A genomic library was constructed in *E. coli* HB101 by the insertion of chromosomal DNA, partially digested with *Sau*3A, into the *Bam*HI site of plasmid pAT153. Three positive clones, shown to direct the synthesis of SpA in *E. coli*, were identified and designated 7-D4, 8-G4 and 10-C9. The recombinant plasmids carried by these clones contained *S. aureus* derived DNA inserts of 8.9 kb, 5.9 kb 6.6 kb in size respectively. Restriction mapping indicated a 3.2 kb *pst*I fragment common to all three plasmids. This fragment was subcloned into the *pst*I site of pBR328 to yield the recombinant plasmid pSPA3. *E. coli* HB101 cells carrying pSPA3 produced two major IgG binding polypeptides with molecular weights of 43 kDa and 41,000 kDa.

Although the above studies demonstrated that *S. aureus spa* genes could be expressed in *E. coli*, the levels of SpA attained (0.5 mg ml^{-1} culture), were of the same order of magnitude as that found in the natural host (Duggleby and Jones, 1983). A primary aim of this study was to elicit the high expression of *spa* in *E. coli*. Such an undertaking would be facilitated by the availability of the entire nucleotide sequence of the *spa* structural gene and its regulatory regions. At the onset of these studies *spa* was known to reside on a 3.2 kb *Pst*I fragment (Duggleby and Jones, 1983). The 5′ end of the gene had been determined by locating the *spa* promoter to a 300 bp region using a beta-galactosidase promoter probe vector. This located the *spa* promoter to a position between 1.9 kb to 2.2 kb from the 5′ end of the 3.2 kb *Pst*I fragment containing the *spa* gene enabling it to be subcloned as a 1.9 kb *Eco*RI/*Pst*I fragment prior to nucleotide sequencing.

3.2 RESULTS

3.2.1 Generation of M13 Template Clones

Nucleotide sequence determination by the dideoxy method (Sanger, *et al.*, 1980) allows up to 350 bp to be accurately read per M13 template on any one gel. Larger regions of

DNA are therefore first fragmented into smaller overlapping segments, inserted into M13 and the sequence data subsequently obtained compiled into a single continuous sequence using computer software. At the inception of this project fragmentation was universally achieved using restriction enzymes. Two strategies could be employed : (i) subcloning using a specific enzyme map; and (ii) the separate "random" cleavage of the DNA with two different restriction enzymes with 4 nucleotide recognition sequences and subsequent insertion of the restriction fragments generated into M13, to yield two distinct template subsets. The restriction enzymes most commonly used in this latter strategy were *Sau*3A and *Taq*I. Both the above strategies were initially used in the sequencing of the SpA gene. The gene had been localised to a 1.9 kb *Eco*RI/*Pst*I restriction fragment. Accordingly, the extremities of this region were isolated as 1.252 kb *Eco*RI/*Hin*dIII and 0.586 kb *Hin*dIII/*Pst*I subfragments and inserted into appropriately cleaved M13mp8 and mp9. In addition, the 1.9 kb *Eco*RI/*Pst*I fragment was isolated by agarose gel electroelution, recovered (2.2.4.6), cleaved with *Sau*3A and *Taq*I, and the cleavage products ligated to M13mp8 digested with *Bam*HI and *Acc*I, respectively. However, upon subsequent transformation of *E. coli* JM101, no recombinant "white" plaques were evident with the latter ligation mix. The suggestion that the 1.9 kb DNA fragment was lacking in *Taq*I sites was confirmed by appropriate digestion of the purified DNA. In contrast, cleavage of the same DNA with *Sau*3A yielded subfragments of 840 bp, 420 bp and 380 bp in size, with evidence of DNA fragments smaller than 150 bp.

To enable the sequence information derived from the above templates to be ordered into a continuous sequence, a further set of recombinant M13 templates was required which carried overlapping subfragments. Rather than utilising a third restriction enzyme with a 4 nucleotide recognition sequence, use was made of a novel procedure (published during the course of this study; Deninger, 1983), which relies on sonication for fragmentation. Accordingly, the purified 1.9 kb *Eco*RI/*Pst*I fragment was subjected to sonication and the sheared fragments blunt-ended by treatment with T4 polymerase.

DNA fragments of between 400 and 800 bp were isolated by size fractionation on agarose gels and inserted into *Sma*I cleaved, dephosphorylated M13mp8.

A summary of the different types of template generated are given in Figure 3.1.

3.2.2 Compilation of Nucleotide Sequence Information

The nucleotide sequence data obtained from the different types of M13 template generated were compiled into a contiguous sequence using a CTL computer and the DNA analysis programmes of Staden (1980). As the sequence data accumulated, a number of problems peculiar to this project became apparent.

The assembly of the sequence data into a complete nucleotide sequence using the two Staden DNA comparison programmes DBCOMP and SEQFIT proved difficult due to the repetitive nature of the *spa* gene in both the IgG binding domains and the cell wall binding region. The programme DBCOMP compares two DNA sequences, usually a new file against the consensus sequence. It allows the minimum match number of sequential bases to be altered, and although it will list the number and position of any matches fitting the set parameters, it will only print the first match found in each contiguous sequence (contig, Staden, 1980) of the consensus sequence, regardless of a better match further in the contig. Therefore, if a newly inputted sequence has two regions that have the same number of sequentially matching bases to the same contig in the consensus sequence, only the first match will be shown. Either the second match has to be assessed by hand or the programme SEQFIT used. SEQFIT will also allow the comparison of a new file against the consensus sequence, however it allows a percentage match parameter to be chosen, rather than a minimum sequential match, and, will display all matches found. A major problem was found with this programme when comparing files against a consensus sequence that contained more than one contig. If sequence data from the new file matched a region of sequence near the end of a consensus contig, the programme would carry the comparison over the into next consensus contig, where very little homology occurred and generate a value for

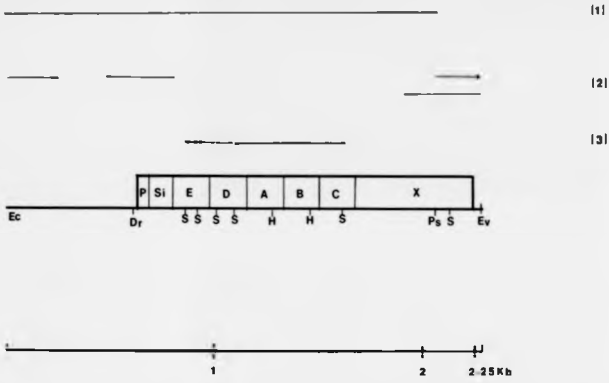


Figure 3.1 Schematic Drawing of the *S. aureus* Cowan I (NCTC8530) *spa* Gene

The figure shows the promoter region (P), signal sequence (Si), IgG-binding domains (E, D, A, B and C) and the cell-wall-and cell-membrane-binding regions (X). The restriction sites used for sequencing are shown; *Eco*RI (Ec); *Dra*I (Dr); *Sau*3A (S), *Hin*dIII (H); *Pst*I (Ps) and *Eco*RV (Ev). The region of the *spa* gene sequenced by each of the different cloning techniques is represented as follows; (1) sonicated DNA, (2) site-directed templates and (3) *Sau* 3A-cleaved DNA.

percentage match which was not necessarily the correct alignment; the position of lower percentage matches within the consensus contigs therefore, had to be taken into account. Normally these limits in the application of the above programmes would not become apparent, as it is unusual to see more than one significant match. The aforementioned problems have occurred entirely due to the high levels of internal homology seen within the *spa* gene.

A similar problem arose when sequencing the cell-wall binding domain (Xr). This consists of a region 255 bp in length and can be interpreted as a set of ten and a half 24 bp repeats, the non identical repeat units vary from one another at only 1 or 2 nucleotide positions (Figure 3.10). Of all the templates derived from sonicated DNA, none were found that contained the entire Xr region. Of the templates analysed that started within the repetitive region it was found that a single dubious base reading was sufficient for the computer to alter the alignment between a new file and the consensus sequence, causing the number of repeat units to vary between 5 and 18. To resolve the number of 24 bp repeats present it proved necessary to run a sequencing gel covering the entire Xr region and at least 24 bp to either side. The only template available that was known to cover this region was a site directed *Pst*I/*Hin*dIII, SpA insert in M13 mp8 which required the region 165 bp to 465 bp upstream of the *Pst*I to be determined. This entailed reducing the ddNTP concentration in the sequencing mixtures which reduced the frequency of dideoxy chain termination and so increased the distance away from the primer to which the sequence could be accurately determined.

A further recurrent problem arose when nucleotide sequence data from *Sau*3A-derived templates was analysed. Many of these templates (greater than 25%) appeared to contain a single consecutive region (i.e., no internal *Sau*3A site) but were found to match two completely separate areas of the 1.9 kb *Eco*RI/*Pst*I fragment. Analysis of the nucleotide sequence around the areas where the mismatch occurred, revealed the presence of one of the following sequences: GAGC, GACC or GATG; leading to the

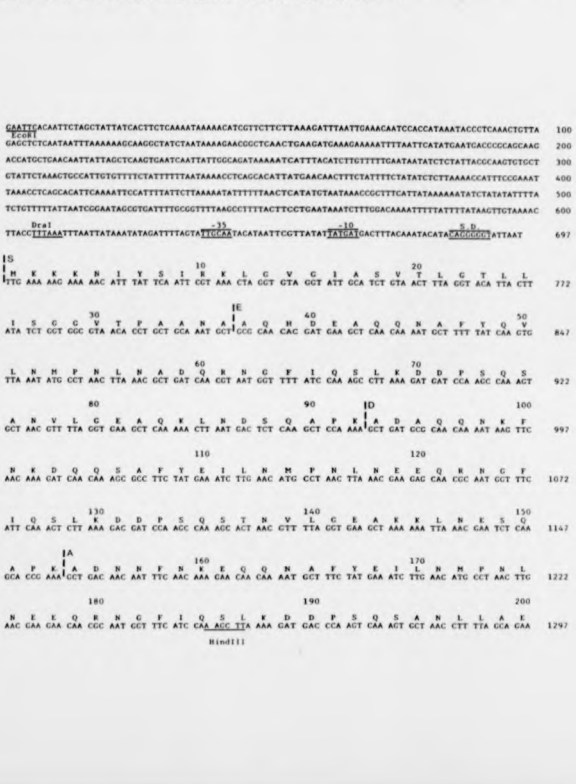
conclusion that the *Sau*3A used was slightly less than specific in its activity, and would recognise a 4 base sequence that only varied from the correct recognition sequence (GATC) for one base at either the third or fourth position. Indeed, later manufacturers instructions from BRL advised that their *Sau*3A was not suitable for sequencing as low levels of non-specific cleavage activity were present.

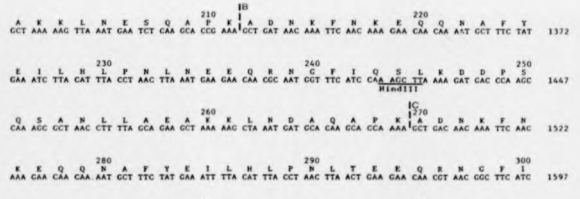
3.2.3 Initial Analysis of the *spa* Nucleotide Sequence

Translation of the DNA sequence into an amino acid sequence revealed an open reading frame 439 codons in length, beginning with a TTG start codon at position 698 (Figure 3.2) and running beyond the *Pst*I site at position 2015. The absence of a termination codon at the 3′ terminus of the 1.9kb *Eco*RI/*Pst*I fragment, indicated that the entire *spa* gene was not carried by pPA3, nor any its derivatives. Plasmid pPA3 directed the expression of IgG-binding polypeptides identical in size to those produced by the clone 8G4 (54 kDa; 43 kDa and 42 kDa), including a band of identical size to native protein A produced by Cowan I (54 kDa). This would appear to indicate that either the cloned *spa* fragment coded for almost the full length protein, or that fusion proteins were being generated between the *spa* gene and the host plasmid at the insertion site. Analysis of the nucleotide sequence of pBR328 and at the site of *spa* DNA insertion, indicated a gene fusion consisting of 438 *spa*-derived codons, followed by a 100 codons (correct reading frame) from the C-terminus of the pBR328 β-lactamase gene. This resulted in a predicted molecular weight for the SpA fusion pre-protein of 61 kDa and SpA fusion mature protein of 57 kDa. In pPA16 the same *spa* codons were fused to 14 codons (incorrect reading frame) of the pUC8 β-galactosidase gene, giving a predicted molecular weight for the SpA fusion pre-protein of 49 kDa and SpA fusion mature protein of 45 kDa. Although the generally accepted molecular weight for protein A, determined by equilibrium sedimentation, varies from 41 kDa to 43 kDa; a molecular weight of 55 kDa to 56 kDa is usually obtained if determined by SDS-polyacrylamide gel electrophoresis (Movitz, 1976. Langone, 1982a).

Figure 3.2 Nucleotide and Predicted Amino Acid Sequence of the *S. aureus* Cowan I (NCTC8530) *spa* Gene

The putative promoter and S.D. sequence are indicated. The signal sequence (S), five IgG-binding domains (E, D, A, B, C) and cell-membrane- and cell-wall-binding region (X, Xr and Xc respectively) are marked.




B
C
X
Xc→
Xc→
HindIII
PstI
EcoRV

3.2.4 Isolation of a Full Length Clone

As neither pPA3 nor pPA16 contained the entire *spa* gene it was necessary to return to the original recombinant plasmids of clones: 8G4, 7D4 and 10C9. Comparison between the restriction endonuclease cleavage map for the pPA3 *spa* fragment (Duggleby and Jones, 1983) and that of the *spa* gene from *S. aureus* 8325-4 (Lofdahl, *et al.*, 1983) revealed 60% homology. The map for 8325-4 extends beyond the *Pst*I site and positions an *Eco*RV site immediately 3′ to the translation termination codon. Therefore the three original clones were cleaved with *Eco*RI and *Eco*RV, all gave a 2.2 kb fragment with restriction sites matching those of pPA3. The 2.2 kb fragment from clone 8G4 was inserted between the *Eco*RI and *Sma*I sites of pUC8 to give pPA30 and transformed into JM83, where it was shown to express low levels of IgG binding polypeptides.

The nucleotide sequence of the 3′ end of the SpA gene from the *PstI/EcoRV* site was obtained by cloning the *Bam*HI-*Hin*dIII *spa* encoding fragment from pPA30 into M13 mp9 and sequencing from the *Bam*HI site across the *Pst*I site towards the *Hin*dIII site in order to overlap the sequence already obtained from pPA3. The sequence of the complementary strand was obtained by cloning the *Bam*HI-*Pst*I *spa* encoding fragment from pPA30 into M13 mp8.

A total of 90% of the entire sequence data (95% of the SpA gene) was derived from both complementary strands, the remaining 10% of sequence data was obtained on only one strand from the site directed templates *Eco*RI/*Hin*dIII and *Pst*I/*Hin*dIII.

3.2.5 Analysis of the SpA Nucleotide Sequence, Predicted Amino-Acid (a.a) Sequence and Secondary Structure

Translation of the DNA sequence into a.a. sequence reveals an open reading frame of 508 codons, beginning at position 698 with a TTG start codon and ending at position 2224 with a TAA stop codon. The protein, including the signal sequence of 36 a.a,

consists of 508 a.a. giving a molecular weight of 55 kDa. The a.a. sequence deduced shows Cowan I SpA to consist of: a 36 a.a. signal sequence; five highly homologous domains (Figure 3.2), 56 to 61 a.a. in length, four of which are known to exhibit IgG binding properties (Sjodahl, 1977b) and a C-terminal region, X, 180 a.a in length which is thought to be responsible for cell wall and cell membrane and cell membrane binding properties of the protein (Sjodahl, 1977a). The X region can be subdivided into a highly repetitive region of 84 a.a. consisting of 10.5 octapeptide repeats (Xr) followed by a non-repetitive region of 81 a.a. (Xc), Figure 3.3.

3.3 DISCUSSION

3.3.1 Codon Usage and Amino-Acid Composition

The genetic code provides two or more options for all the amino acids except methionine and tryptophan which are coded for by a single codon. From the sequence studies carried out on many procaryotic and eucaryotic mRNA species, it is evident that preferences exist for the use of alternative codons; both between different species and between the mRNA involved in normally highly expressed genes compared with that involved in weakly expressed genes.

It is thought that the non-random distribution of synonymous codons may be related to the level of gene expression, by affecting the efficiency of translation. Examination of the correlation in *E. coli* codon usage and content of the respective tRNA by Ikemura (1981a and 1981b) and Ikemura and Ozeki (1982), together with studies to determine which synonymous codon is preferred in strongly and in weakly expressed genes (Gouy and Gautier, 1982; Grojean and Fiers, 1982; Konigsberg and Godson, 1983; and Wain- Hobson *et al.*, 1981), has enabled the construction of a set of proposals to explain the codon preference exhibited. These proposals can be subdivided into two groups; those dependent on the levels of the various tRNAs available to the gene, and those concerned with codon-anticodon interaction energies.

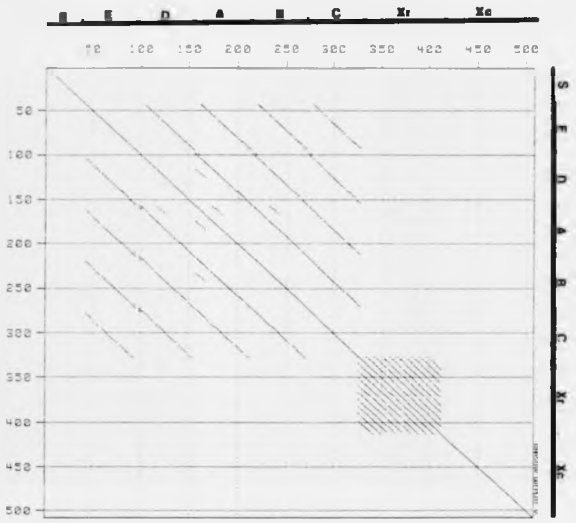


Figure 3.3 A dot Matrix Homology Plot for SpA (NCTC8530)

The plot shows areas of internal homology within SpA. The axes relates to the amino acid sequence of SpA (with position 1 on each axis corresponding to the N-terminal residue of the signal peptide). The functional domains of SpA are indicated parallel with the axes. The plot shows two regions to be homologous if six or more residues from a block of ten are identical (and in the same relative position) within the block.

The most important appears to be the correlation, reported by Ikemura (1981a), between highly abundant tRNAs and a strong bias towards their respective codons in highly expressed genes. Ikemura and Ozeki, 1982, also state that, where two isoacceptors are present in equal amounts in the cell, one of which responds to two codons whereas the other responds to only one of the two codons, then the codon that is recognised by both the tRNA isoacceptors will be more frequently used than the other. It is proposed that codons corresponding to minor tRNAs or weak isoenzymes, may act as modulator codons regulating the elongation rate of the nascent polypeptide during translation.

The second group of rules concerning the choice between synonymous codons appears to be on the basis of the optimisation of codon-anticodon pairing energies, Grosjean and Fiers (1982), have suggested that the translational efficiency is facilitated by intermediate strength reactions rather than strong or weak reactions. In general, codons of the (A/U)-(A/U)-pyrimidine type form intrinsically weak interactions, preferring a C in the wobble position to stabilise the codon- anticodon interaction. Codons of the (G/C)-(G/C)- pyrimidine type are involved in intrinsically strong interactions and prefer a U in the wobble position, to weaken the codon- anticodon interaction and so optimise translation. In certain circumstances, such as modification of the nucleotide in the wobble position, the above rules have been found to be invalid. A number of such examples have been reported (Ikemura, 1981a and 1981b; Ikemura and Ozeki, 1982). Thiolation of uridine, such as that found in *E. coli* $tRNA_2^{Glu}$ and $tRNA^{Lys}$, or 5′-carboxymethalation of uridine restricts wobble G.U pairing and so produces a preference for using the A-terminated codon over the G-terminated codon. Inosine in the wobble position produces a preference for U and C terminated codons over the A terminated codon and so leads to purine-purine pairing. It has also been noted by Gouy and Gautier, 1982; and Nichols *et al.*, 1980) that the aforementioned pyrimidine-ending codon preference is only independent of amino acid composition in highly expressed genes. In weakly expressed genes codon usage appears to more consistent with the G+C content of the genome, and where the G+C content is

unusually high or low this bias is often disproportionately represented in the wobble position of the codon.

3.3.1.1 Codon usage in Staphylococcus species

Unlike *E. coli* no major studies have been undertaken on *S. aureus* or other staphylococcal species to establish the number of tRNA species and their abundance within the cell. Although the studies on the closely related enteric organisms *E. coli* and *Salmonella typhimurium* (Ikemura, 1985) have shown a similar pattern of codon usage, *S. aureus* is a Gram-positive organism unrelated to the former two organisms and has a different G+C content, and thus might exhibit a different codon usage pattern. From the fact that organisms as diverse as *Saccharomyces cerevisiae* and *E. coli* have shown that their genes differ more in the degree of bias rather than its direction, it might be predicted that other procaryotes would have a codon usage pattern similar to that of *E. coli*. However the only Gram-positive organism which has been examined for codon usage preference to any great extent, *B. subtilis*, (examination of the nucleotide sequences of 21 *B. subtilis* genes by Ogasawara (1985) and 56 *B.subtilis* genes by Shields and Sharp (1987)) has revealed several differences in codon usage from *E. coli*. Firstly, there appears to be comparatively little codon bias in *B.subtilis* genes, substantial codon bias only being seen in very highly expressed genes. Secondly, there is no apparent preference for codons with an intermediate codon-anticodon interaction energy. Finally there appears to be no evidence of selection in favour of purine- any nucleotide-pyrimidine codons (G/A-X-C/U), reported by Shepherd (1984) to be seen in excess in most other organisms. It is has been proposed by Vold (1985) that the unique organisation of tRNA genes of *B. subtilus* is responsible for the unusually unbiased codon usage of this organism. As the tRNAs present in *S. aureus* have not been recorded it it not possible to predict codon usage for S.aureus or other staphylococcal species. Therefore codon bias will have to be determined by the examination of the known nucleotide sequences for staphylococcal genes. This data is very limited and unfortunately, most of those sequenced to date reside on staphylococcal plasmids and transposons, and as such may have originated from a different bacterial

species. A comparison between the codon usage predicted from the *spa* nucleotide sequence and the codon usage predicted from the nucleotide sequences of 15 staphylococcal genes (further details of genes selected are given in Table 3.1), is presented in Table 3.2. A substantial difference between the G+C content of the plasmid derived genes and the chromosomal and transposon genes is evident (Table 3.2), a factor which may influence the codon bias (as stated earlier). Accordingly, the staphylococcal genes were segregated into the three different categories, on the basis of origin, (Table 3.2), and three separate comparisons undertaken.

The comparative date obtained revealed slight differences in the codon preference between the 3 categories (chromosome, plasmid and transposon) of genes examined: plasmid genes had a higher content of the codons GCU^{Ala} and CAA^{Gln}; chromosomal genes had a higher content of the codons CGU^{Arg}, GGU^{Gly} and AAG^{Lys}, and a lower content of the codon ACU^{Thr}; while transposon genes had a higher content of the codon UGC^{Cys}. Comparison of *spa* with these preferences revealed that it showed greater similarity to plasmid codon preference for residues Ala, Gln and Lys, with only the residue Cys showing the same codon preference as the chromosomal genes.

The *spa* nucleotide sequence showed a codon bias substantially different to all the other 15 staphylococcal genes examined for eight particular residues. 1) Of the six codons corresponding to the residue Arg, there was a marked preference for the codon CGC (44% of the Arg residues), while AGA was never used. In the case of the other 15 genes, AGA and CGC accounted for between 35%-65% and 0%-3% respectively, of the Arg codons. 2) Of the two codons corresponding to the residue Glu, GAA was seen to be approximately 25% higher and GAG 25% lower than the corresponding codons in the other 15 genes. 3) Of the four codons corresponding to Gly, GGA was seen to account for only 8% of the residues, rather than 26%-39% in the other staphylococcal genes. 4) Of the six codons corresponding to the residue Ser, there is a

TABLE 3.1

Staphylococcal Species Genes Selected for Codon Usage Comparison

Gene/Product of DNA	Organism content	Origin reference	%G+C	Sequence
Enterotoxin B	*S. aureus*	C	37.9	Jones & Khan, 1986
Staphylokinase	*S. aureus*	C	34.5	Sako & Tsuchida, 1983
Lipase	*S. hyicus*	C	39.5	Gotz *et al.*, 1985
Alpha toxin	*S. aureus*	C	34.5	Grey & Kehoe, 1984
Protein A	*S. aureus*	C	38.1	Shuttleworth *et al.*, 1987
Exfoliative toxin B	*S. aureus*	P	28.7	Jackson & Iandolo 1986
Lysostaphin	*S. simulans*	P	37.1	Recsei *et al.*, 1987
*lin*A &(Lincomycin R)	*S. haemolyticus*	P	28.6	Brisson-Noel Courvalin, 1986
β-lactamase	*S. aureus*	P	27.9	Wang & Novick, 1987
spc (Spectinomycin R)	*S. aureus*	T	36.4	Murphy *et al.*, 1985
*erm*A (Erythromycin R)	*S. aureus*	T	32.5	"
*tnp*A (transposition)	*S. aureus*	T	32.2	"
*tnp*B (transposition)	*S. aureus*	T	33.9	"
*ynp*C (transposition)	*S. aureus*	T	39.9	"

* Organism from which DNA was isolated

Origin of DNA: C = chromosome, P = plasmid and T = transposon

All the above genes are thought to be moderately expressed excepting B-lactamase and Protein A which are more strongly expressed and the transposition genes which are weakly expressed.

TABLE 3.2

Codon usage in 15 staphylococcal species genes compared with that for *spa*

Residue and codon	Source of gene							
	Plasmid		Chromosome		Transposon		SPA	
	A	%B	A	%B	A	%B	A	%B
Ala GCA	17	27	35	48	28	41	17	34
Ala GCC	8	13	8	11	8	12	1	2
Ala GCG	1	2	16	22	8	12	4	8
Ala GCU	36	58	14	19	24	35	28	56
Arg AGA	15	65	14	35	50	54	0	0
Arg AGG	0	0	3	8	15	16	0	0
Arg CGA	3	13	3	8	5	5	0	0
Arg CGC	0	0	1	2	3	3	4	44
Arg CGG	1	4	3	8	4	4	0	0
Arg CGU	4	18	16	40	15	16	5	56
Asn AAC	14	17	18	26	23	21	42	70
Asn AAU	67	83	50	74	86	79	18	30
Asp GAC	15	23	19	21	11	12	17	45
Asp GAU	51	77	72	79	80	88	21	55
Cys UGC	0	0	0	0	5	29	0	0
Cys UGU	4	100	1	100	12	71	0	0
Gln CAA	31	97	38	83	54	70	37	100
Gln CAG	1	3	8	17	23	30	0	0
Glu GAA	51	77	67	77	89	64	39	95
Glu GAG	15	23	20	23	50	36	2	5
Gly GGA	27	39	23	26	34	44	3	8
Gly GGC	5	7	19	21	6	8	13	33
Gly GGG	4	6	12	13	11	14	0	0
Gly GGU	33	48	36	40	27	35	24	60
His CAC	7	26	13	30	10	22	1	20
His CAU	20	74	30	70	35	78	4	80
Ile AUA	29	39	23	32	49	37	1	6
Ile AUC	5	7	12	16	14	11	9	50
Ile AUU	41	55	38	52	70	52	8	44

Table 3.2 continued

Residue and codon	Source of gene: Plasmid A	Plasmid %B	Chromosome A	Chromosome %B	Transposon A	Transposon %B	SPA A	SPA %B
Leu CUA	7	9	8	11	28	15	4	9
Leu CUC	1	1	0	0	7	4	0	0
Leu CUG	3	4	5	7	9	5	0	0
Leu CUU	9	12	9	13	35	19	8	19
Leu UUA	47	60	33	46	67	36	27	63
Leu UUG	11	14	16	23	38	21	4	9
Lys AAA	96	79	19	58	127	70	51	82
Lys AAG	25	21	14	42	54	30	11	18
Met AUG	24		24		41		4	
Phe UUC	8	22	10	24	20	25	13	81
Phe UUU	29	78	31	76	60	75	3	19
Pro CCA	20	51	19	40	28	46	10	32
Pro CCC	1	3	1	2	4	7	0	0
Pro CCG	3	8	9	18	4	7	2	6
Pro CCU	15	38	19	40	24	40	19	62
Ser AGC	9	12	14	15	12	11	11	48
Ser AGU	17	22	14	15	32	29	4	17
Ser UCA	31	41	31	33	27	24	3	13
Ser UCC	0	0	5	5	7	6	0	0
Ser UCG	3	4	12	13	7	6	0	0
Ser UCU	16	21	19	20	27	24	5	22
Ter UAA	2	100	3	100	2	50	4	67
Ter UAG	0	0	0	0	0	0	1	17
Ter UGA	0	0	0	0	2	50	1	17
Thr ACA	34	46	61	56	35	38	4	36
Thr ACG	2	3	11	10	14	15	0	0
Thr ACG	7	9	19	17	15	16	0	0
Thr ACU	33	43	17	16	29	31	7	64
Trp UGG	13		16		25		0	
Tyr UAC	13	21	10	21	8	11	0	0
Tyr UAU	49	79	37	79	63	89	6	100
Val GUA	36	47	21	30	39	30	6	43
Val GUC	4	5	14	20	17	13	2	7
Val GUG	6	8	10	14	21	16	2	14
Val GUU	30	39	25	36	51	40	5	36

Columns headed A represent total number of codons. The codon usage was derived from the nucleotide sequences of the genes present in Table 3.1.

Columns headed %B represent percentage synonym use, and was obtained by dividing the number of times a codon was used in the relevant genes by the number of times all codons specifying the same amino acid were used , expressed as a percentage.

Ter, Termination codon.

minimum of 69% higher content of the codon AGC and a significantly lower content of the codons UCA and UCC, compared to that found in the other staphylococcal genes. 5) The codon AUC, corresponding to the residue Ile, is substantially higher in the *spa* gene while the synonymous codon AUA is greatly reduced. 6) There is little codon bias shown between the two synonymous codons corresponding to the residue Asp in *spa*, whereas in the other staphylococcal genes (excepting *erm*, equal codon usage) the codon GAU is far more common than GAC. The most substantial difference in codon bias is seen for the residues Asn and Phe where 7) the codon content between the two possible synonyms for Asn in *spa* are 70% (AAC) and 30% (AAU), whereas completely the opposite usage is seen in the other staphyloccocal genes 17-26% (AAC) and 74-83% (AAU), except in *erm* where AAC represents 45% of the codon content and AAU 55%. 8) An almost identical bias is seen for the residue Phe, with codon usage in *spa* being 81% (UUC) and 19% (UUU), compared with 22-25% (UUC) and 75-78% (UUU)in the other staphylococcal genes except that for alpha toxin where codon usage is 50% of each synonym.

The most likely explanation for the difference in codon bias between the *spa* sequence and the sequences of the other staphylococcal species genes examined, is the fact that SpA is a fairly abundant cell protein (1.7% of total cell protein, Sjoquist *et al.*, 1972b) whereas all the other genes examined code for less abundant proteins. The codon bias seen in *spa* may be indicative of the codon bias needed for near to optimum translation potential in *S. aureus*. All the above codon preferences (points 1 to 8) exhibited by the *spa* sequence correspond to the codon bias present in strongly expressed *E. coli* genes (Table 3.3). This would appear to indicate that the codon usage pattern of *S. aureus* is very similar to that of *E. coli*, however until further *S. aureus* strongly expressed genes have been sequenced, their codon usage determined and tRNA abundance in *S. aureus* examined the above statement cannot be substantiated.

TABLE 3.3

Codon Usage in *spa* Compared with Optimum Codon Usage in *E.coli*

Residue and codon	optimal tRNA abundance in E coli	optimal codon -anticodon binding strength	codon usage in strongly expressed E coli genes	† spa syn. use %	* E coli syn. use %
Ala GCA	O			34	23
Ala GCC	N	N		[illegible]	19
Ala GCG	O			[illegible]	30
Ala GCU	O	O		56	28
Arg AGA	N		I	0	1
Arg AGG	N		R	0	[illegible]
Arg CGA	O		I	0	[illegible]
Arg CGC	O	N		44	35
Arg CGG	N		I	0	3
Arg CGU	O	O		56	58
Asn AAC	O	O		70	76
Asn AAU	O	N	R	30	24
Asp GAC	N			45	49
Asp GAU	N			55	51
Cys UGC	N			0	58
Cys UGU	N			0	42
Gln CAA	N		R	100	27
Gln CAG	O			0	73
Glu GAA	O	O		95	73
Glu GAG	N	N	I	5	27
Gly GGA	N		I	8	5
Gly GGC	O	N		33	41
Gly GGG	N		I	0	7
Gly GGU	O	O		60	48
His CAC	N			20	61
His CAU	N			80	39
Ile AUA	N		R	[illegible]	[illegible]
Ile AUC	O	O		50	62
Ile AUU	O	N		44	37
Leu CUA	N		I	9	2
Leu CUC	N		I	0	7
Leu CUG	O			0	69
Leu CUU	N		I	19	9
Leu UUA	N		I	63	6
Leu UUG	N		I	9	8
Lys AAA	O	O		82	77
Lys AAG	O	N	I	18	23

Residue and codon	optimal tRNA abundance in E coli	optimal codon -anticodon binding strength	codon usage in strongly expressed E coli genes	†spa syn. use %	*E coli syn. use %
Met AUG	N			-	-
Phe UUC	N	O		81	56
Phe UUU	N	N		19	44
Pro CCA	O			32	20
Pro CCC	N	N	R	0	6
Pro CCG	O			6	65
Pro CCU	N	O	R	62	9
Ser AGC	N			48	22
Ser AGU	N		I	17	6
Ser UCA	N		I	13	8
Ser UCC	N			0	26
Ser UCG	N		R	0	11
Ser UCU	N			22	27
Thr ACA	N		R	36	6
Thr ACC	O			0	51
Thr ACG	N		R	0	20
Thr ACU	O			64	24
Trp UGG	N			0	-
Tyr UAC	O	O		0	59
Tyr UAU	O	N		100	41
Val GUA	O			43	23
Val GUC	N			7	13
Val GUG	O			14	27
Val GUU	O			36	38

O: optimal codon. N: non-optimal codon. Determination of optimal and non optimal codons correlating to tRNA abundance was according to Ikemura (1981), and Ikemura and Ozeki (1982). Determination of optimal and non-optimal codons correlating to optimal codon-anticodon binding interaction was according to Grosjean and Fiers (1982), Ikemura (1981) and Ikemura and Ozeki (1982).

I: Infrequently used codons. R: Rarely used codons. Determination of infrequently and rarely used codons was according to Konigsberg and Godson (1983); where, of the 25 *E.coli* genes examined, infrequently used codons totalled an average of 12.1% of all codons used, and rarely used codons totalled an average of 4.2% of all codons used.

†*spa* and **E. coli* percentage synonym (syn.) usage was calculated as detailed in the legend to Table 3.2. Codon usage for the *spa* gene is also detailed in Table 3.2 and codon usage for the 25 *E.coli* genes can be found in the paper by Konigsberg and Godson, 1983.

3.3.1.2 Expression potential of *spa* in *E. coli* indicated by codon bias

One of the main objectives of this project was to synthesise SpA in *E. coli* at a higher level than by the native organism. As translational efficiency is a major factor in final protein levels, a comparison of codon usage of the *spa* gene against that preferred by strongly expressed *E. coli* genes and the relative abundance of the corresponding *E. coli* tRNA species was performed; the results are presented in Table 3.3.

The *spa* gene appears to have the same codon bias for many of its residues (Lys, Ser, Val, Arg, Asn, Asp, Gly, and Ile) as seen in most highly expressed *E. coli* genes. For the residues Ala, Glu and Phe the direction of codon bias is the same as that for highly expressed *E. coli* genes, however, the degree of bias is greater. Several SpA residues; Gln, His, Leu, Pro, Thr and Tyr have a completely different codon preference to that of *E. coli*. The codon bias seen for these six residues in *spa* is apparent throughout all staphylococcal genes sequenced to date (see Table 3.2). With the exception of Tyr (where the optimum synonym depends on codon- anticodon interaction strength), all have the synonym bias, which in *E. coli* has been shown to be dependent on tRNA abundance. Such a strong codon preference throughout all the staphylococcal species genes examined may be indicative of a difference between *E. coli* and staphylococci in the relative abundance of the different tRNA species. The presence of a different synonym bias for Gln, His, Pro and Thr will probably not reduce the expression potential of the *spa* gene in *E. coli*, to any great extent, as there is only a minor difference between the *E. coli* cellular content of the tRNA species corresponding to the preferred synonym in *spa* and that preferred by *E. coli*. Whereas for the residue Leu there is a major difference in their relative tRNA abundances, with the *E. coli* content of the tRNA corresponding to the preferred *E. coli* synonym for Leu (CUG) being 4 times higher than the tRNA content corresponding to UUA (the preferred Leu synonym in *spa*, (Ikemura, 1981a&b). Whether the predominant use of the synonym UUA in the *spa* gene will result in a reduction of its translation rate in *E. coli* may well depend on their position within the gene. Robinson *et al.*, (1984) inserted a 6 amino acid synthetic oligonucleotide into the CAT (chloramphenicol acetyltransferase) gene of *E.*

coli such that a favoured synonym was substituted with a rarely used synonym for four consecutive residues. This modification had no effect on the amount of protein produced at moderate transcription rates, however, upon maximal induction of the preceding *trp* promoter the synthesis of CAT and the CAT-derived protein from the gene with the favourable synonym insert increased 3-fold, whereas the synthesis of the CAT-derived protein from the gene with the rare synonym insert did not increase.

The position of the synonyms corresponding to Leu in the *spa* gene are shown in Figure 3.4. Although the close proximity of UUA codons in the N-terminal region of IgG-binding regions B and C may have some effect on the rate of translation, it is more likely that the cell-membrane association region, Xc, may limit the translation rate of the *spa* gene in *E. coli*. This hypothesis is corroborated by later experiments (section 4.2.1.2) where a plasmid (pPA34) carrying a C-terminally truncated *spa* gene, lacking the Xc region, produces almost twice as much SPA as a similar plasmid (pPA31) carrying the entire *spa* gene.

The most noticable difference in codon usage between the *E. coli* and the staphylococcal species genes examined is in the choice of translation initiation codon. The codons AUG, GUG and UUG are thought to be able to function as initiation codons as they have all been shown to stabilise the binding of initiator fMET-tRNA to ribosomes *in vitro* (Clark and Marcker, 1966). Comparison of the initiation codons of 126 *E. coli* genes (Gen Bank, 1984) have shown a marked preference for ATG to be present as the initiation codon, 115 (91.2%) times against 6 (4.5%) TTG initiation starts and 5 (4.0%) GTG initiation starts. Comparison of the initiation codons from 20 staphylococal species genes (DNASTAR Gen bank, 1987) shows a less marked preference for ATG as the initiation codon, 13 (65%) ATG initiation starts against 4 (20%) TTG initiation starts and 1 (5%) GTG initiation starts. Reddy *et al.*, (1985) has shown that replacement of the TTG initiation codon in the *E. coli* adenylate cyclase (*cya*) gene with ATG and GTG resulted in a 3-fold and 2-fold increase in the gene

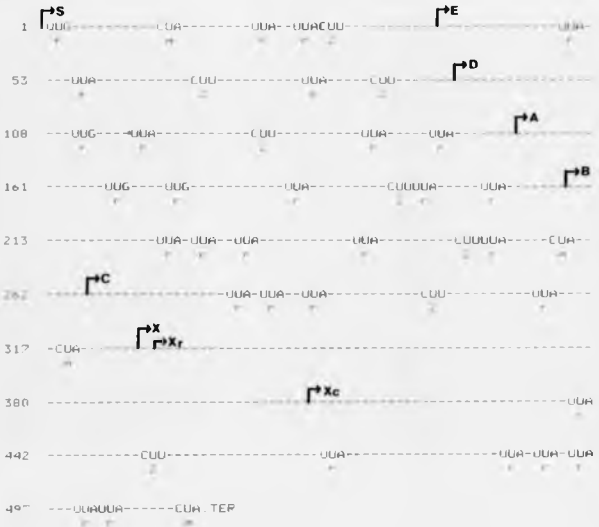


Figure 3.4 Position of Synonymous Codons Corresponding to Leucine Residues in *spa*

Amino acid sequence positions are indicated down the left hand side of figure, the different domains of SpA are indicated above the sequence and are designated as already described in Figure 3.2. Codons are represented by a -, except for the synonymous codons corresponding to Leucine which are written in full. Lettering beneath the codon designates the tRNA which corresponds to that particular codon (nomenclature is that used by Ikemura, 1981); r:*Leu* tRNA UUR (*E. coli* content = 0.24), m : *Leu* tRNA CUA (*E. coli* content = minor), 2 : *Leu* tRNA CUU (*E. coli* content = 0.3).

product, respectively. Thus although it is not unusual for *spa* to possess a TTG initiation codon it might adversely affect its translation initiation frequency when expressed in *E. coli*.

3.3.2 Features of the 5′ Non-Coding Regions

3.3.2.1 Transcription Initiation Signals

A promoter is defined as a segment of DNA that contains signals for the proper binding and subsequent activation of RNA polymerase holoenzyme, to a form capable of initiating the synthesis of RNA, it usually precedes the coding section of the gene.

Comparison of the nucleotide sequences of *E. coli* and related bacteria (Raibaud and Schwartz, 1984; Rosenburg and Court; 1979) have revealed several common features, enabling the construction of a consensus sequence for the general prokaryotic promoter. Numbering of the DNA sequence of this region is such that nucleotide +1 is the start of the mRNA transcript. There are two conserved hexanucleotide sequences: one around the -35 position consisting of a highly conserved trinucleotide TTG and a less conserved ACA sequence, (TTGaca); the other around the -10 position, often referred to as the Pribnow box (Pribnow, 1975), consists of three highly conserved nucleotides in the first, second and sixth positions and three less conserved nucleotides in the centre, TAtaaT. The -35 promoter sequence is thought to be important for the initial recognition of the transcription initiation region by RNA polymerase (Schaller, 1975). The -10 promoter sequence is proposed as the initial binding site for RNA polymerase at which the helical DNA unwinds in order for initiation to occur at a purine residue normally 7-9 bp downstream (Rosenberg and Court, 1979).

The Cowan I *spa* nucleotide sequence (Figure 3.3) contains 2 hexanucleotide sequences that correspond closely to the consensus -35 and -10 promoters. At positions 638 to 643 the *spa* nucleotide sequence TTGCAA only differs from the -35 consensus sequence in the less conserved trinucleotide region by an inversion of the fourth and fifth bases. At positions 661 to 666 the *spa* sequence TATGAT differs from the -10

consensus sequence at the less highly conserved fourth nucleotide where a G (instead of an A) is present. Further evidence in support of these two hexanucleotides representing the position of the transcription initiation region of the *spa* gene is obtained by comparison with other procaryotic promoters. Firstly, the -10 promoter is in close proximity, 14 bp upstream, to a possible ribosome binding sequence (see section 3.3.2.2). Secondly, the nucleotide spacing between the -35 and -10 promoter has been shown by Stefano and Gralla (1982a) to be a crucial factor in promoter strength. Almost all the prokaryotic promoters they examined had a spacing of 17 $\pm$ 1 with 17 considered to result in maximum promoter strength. Thus, the SpA promoter region is not only typical, with a spacing of 17 nucleotides, but is also potentially a very strong promoter. Thirdly, the rigidity of the DNA in the promoter region has been proposed to play a crucial role in its interaction with the RNA polymerase (Travers, 1986). The rigidity of DNA is often dependent upon the presence of rigid homopolymer dA:dT stretches, usually 4 bp or greater in length. The statistical survey undertaken by Travers (1986), indicated that such rigid sequences occur most frequently around the -45 and -55 position and are preferentially excluded from the regions immediately upstream of the -35 and -10 sequences, and from a region around -23 to -25. The nucleotide sequence of the proposed *spa* promoter region conforms to this pattern in that it has no A or T blocks greater than two nucleotides in the region from the -35 to the -10 promoter and has a stretch of 4 T's from positions -43 to -46.

3.3.2.2 Translation initiation region

In order for translational initiation to occur, a small ribosome subunit must specifically bind to the region of the mRNA containing the initiation codon. Hence there must be some mode of differentiation by the ribosome subunit between an initiation codon and internal AUG, GUG and UUG codons. Shine and Dalgarno (1975) first proposed the presence of a specific base sequence prior to the initiation codon, after observing that all of the coliphage RNA ribosome-binding sites reported and the ribosome- binding site of an endogenous *E. coli* mRNA, contained all or part of the sequence 5' -AGGAGGU-3' (Shine-Dalgarno sequence, S.D), which in turn is complementary to

the sequence (5′ -ACCUCCU-3′) they had shown to be present at the 3′ terminus of *E. coli* 16S RNA. This would suggest a direct base pairing role in the initiation of protein synthesis, rather than a dependence on the secondary structure of the region.

Subsequent analysis of the translation initiation region (-25 to +20, where position 1 represents the first nucleotide of the initiation codon) from 123 DNA sequences by Stormo *et al.*, (1982) revealed a number of additional features present in the preferred ribosome binding region. There appeared to be discrimination against the purine G occuring on both sides of the S.D sequence, possibly to diminish misalignment during annealing to the 16S RNA. A preference for the pyrimidine T rather than C, which may prevent complementation with the purine G tract of the S.D sequence, and a preference for the purine A in most of the region outside the S.D sequence itself. A further feature noted was a translation termination codon commonly found immediately 5′ to the initiation codon. Atkins (1979) has examined this latter point in detail and reports that in 50% of the 92 sequences considered, a termination codon occurs within the first three codons 5 to the translational start and in nearly all the remainder within the first seven codons. It was also noted that 60% of the sequences have a termination codon (in a different reading frame to that of the gene) within the first four codons following the initiation codon. The proportion of each of the termination codons is noticeably different, with TAA occuring twice as often as TGA, and TAG not present as the first termination codon preceding the translation initiation codon in any of the functional genes examined. Furthermore, TAG only constitutes 5% of the termination codons found immediately following the initiation codon. Atkins has proposed that the termination codons TAA and TGA may be very important for ribosome recognition. However, a later review by Gold *et al.*, (1981) concludes that although the frequencies of the different termination codons reported by Atkins are correct, it is purely coincidental that they are termination codons. Their presence can be explained as a side effect of the preference for the purine A and the discrimination against G and C seen in this region. Gold *et al.* (1981) also points out that the non-termination codon TTA is present at the same level as TAA. Also the apparently

significant discrimination against TAG is due, not only to the fact that it is the least abundant of all the trinucleotides, but also because both TAA and TGA are complementary to the 3′ end of the 16S RNA (3′-AUUCCUCCACU-5′) and can overlap the initiation codon, whereas TAG cannot. Although Gold *et al.* (1981) queries the role of termination codons in translation initiation, he accepts that there is evidence that the ribosomes might scan for T's and A's in this region, in particular the sequence TTAA.

The nucleotide sequence immediately preceeding the *spa* structural gene complies extensively with the aforementioned characteristics of a translation initiation region. Within the initiation region there are only two C's 5′ to the S.D sequence and none 3′ to it. Furthermore, apart from the S.D sequence, this region contains only two G's (one at position +3 and the other at position +9). The entire region is very A rich. Thus, excluding the poly G tract of the S.D sequence, 23 of the 40 nucleotides are A's. The "termination" codons TAA and TGA are found respectively, 4 nucleotides prior to, and overlapping, the initiation codon. The actual *spa* S.D sequence (5′-AGGGGGT-3′) only varies at one nucleotide position, the central G, from the accepted consensus sequence (5′-AGGAGGT-3′) and thus would form a stable ribosome binding complex with a ΔG of -12.2 Kcal mol^{-1}. The number of nucleotides between the S.D sequence and the initiation codon in *spa* is 6, which falls within the normal range of 5 to 9 .

Unlike the RNA polymerase binding site, variations in both the nucleotide spacing between pertinent features within the translation initiation region and the large differences in free energy values (-4 to -22 Kcal mol^{-1}, Stormo 1982) determined for ribosome binding, appear to have no correlation with translation initiation rates. Experimental evidence reviewed by Gold *et al.* (1981) indicates that the secondary structure, in particular hairpin loops, of this region may play an important role in the control of initiation, with rates dependent upon the accessibility of the S.D sequence and the initiation codon. The nucleotide sequence of the *spa* gene gives no evidence of any such secondary structure control mechanism in the translation initiation region, though it is possible that tertiary structure control mechanisms may exist.

3.3.2.3 Features 5′ to the *spa* -35 transcription initiation region

The regulation of gene activity is often controlled by the regulation of the frequency of transcription initiation. As the understanding of the promoter regions of genes has increased, it has become apparent that initiation of transcription in many systems is strongly influenced by sequences well upstream of the genes they regulate i.e., *lac* operon, *gal* operon, *ara* operon, the genes coding for tryptophanase, D-serine cheaminase and histidase are just a few examples.

These initiation control signals can be divided into two categories, (i) those where the actions of regulatory proteins inhibit transcription initiation (repressors), and (ii) those where the action of regulatory proteins stimulate transcription initiation (activators). These regulatory proteins are thought to operate to enhance or prevent the binding of RNA polymerase, by one of two modes (i) interaction with the RNA polymerase itself or (ii) interaction with the nucleotide sequence surrounding a potential RNA polymerase binding site. Such interaction may cause an alteration in the DNA configuration which will effect the binding rate of the RNA polymerase i.e., it may enhance a sequence that normally varies from the recognised consensus to such an extent that it is not recognised by RNA polymerase; the presence of an activator may allow recognition or alternatively may alter the conformation of a previously inhibitory nearby DNA sequence allowing promoter activity. One of the most common regulator proteins known to be an activator for the expression of *lac* and other systems controlled by catabolite repression is the catabolite receptor protein (CRP), the product of the *crp* gene. In the presence of 3′-5′ cyclic adenosine monophosphate (cAMP), the protein has been shown to bind at specific sites in the vicinity of several promoters and have a stimulatory effect on transcription initiation. Examination of the CRP binding site (CBS) of several genes (Ebright, 1982; Raibaud and Schwartz, 1984) has shown that 10 nucleotides within a continuous 17 nucleotide sequence are normally conserved, Ebright proposes that these nucleotides constitute the CRP-DNA contact points. Computer analysis of further 7 operator regions from genes known to have CRP-

binding sites by Studnicka (1987) has shown that some nucleotides of the binding site are more strongly conserved that others, those strongly conserved are shown below.

The nucleotide sequence upstream of the *spa*-10 promoter was examined for homology with CBS and other known operator sequences. Two possible CBS/operator sites (CBS1 and CBS2) were identified (Figure 3.5) between nucleotides 93 and 117 and the second between nucleotides 292 and 316, these are shown below compared with CBS consensus.

CRP binding site sequence	a a . T G T G A . . t a . . t C a a/c A t a/t t . a/t g .	
CBS1	A A c T G T t A g a g c t c T C A A t a A T t T a	76% homology
CBS2	A A g T G c t g T A t t c t A A A g T g c c a	72% homology

spa CBS1 homology has a spacing of 6 nucleotides between conserved regions and also shows a region of dyad symmetry associated with the 3′ and 5′ halves of the CBS or operator site. The *spa* CBS2 homology has a spacing of 5 nucleotides between the two conserved regions, though an extra spacing of one nucleotide would also realign the *spa* CBS to increase the level of homology among the most strongly conserved nucleotides of the 3′ half of the binding site.

Abrahmsen *et al.*, 1986 reported experimental results following "heat-shock" treatment that suggest the existence of a stress induced promoter in front of the *spa* gene. Comparison of the region upstream of the -35 *spa* promoter with known "heat shock" promoters (Reznikoff *et al.*, 1985) proposes putative -35 and -10 sites (shown in Figure 3.5).

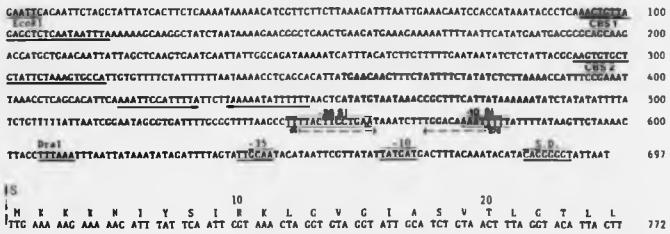


Figure 3.5 **Features 5′ to the *spa* Coding Region**

The nucleotide sequence upstream of the none heat shock -10 promoter was also examined for sequences that could result in a secondary structure formation which would affect transcription from a cloning vector promoter positioned upstream of the *Eco*RI site. Only one small area of dyad symmetry was found (the position of which is indicated in Figure 3.5) this region was short in length (31 bp) and apart from the potential to form an RNA hairpin loop it exhibits none of the structural features (such as a G.G inverted repeat sequence followed by a run of consecutive thymidine residues in the non template DNA strand), normally associated with termination sequences. This region is unlikely to have any effect other than decreasing the rate of transcription from an upstream vector promoter.

Although the *spa* gene is known to be subject to regulation in *S. aureus* this is dependent on growth phase of the cell (production has been shown to cease after stationary phase has been reached Lindmark *et al.*, 1977); not on catabolite levels. The *agr*A gene (accessory gene regulator) and *agr*B and *agr*C genes together with the associated δ-lysin gene (*hld*) have been shown to be involved in the *agr* locus regulation of post-exponential-phase synthesis of staphylococcal proteins including SpA (Recsei *et al.*, 1986; Peng *et al.*, 1988; Janzon *et al.*, 1986, 1989; Novick *et al.*, 1989 and Janzon and Arvidson, 1990). The *agr*A gene is thought to act at the mRNA level, though its exact mode of action is unknown. All known regulatory RNAs in procaryotic systems seem to be antisense RNA which act as transcription terminators or translation inhibitors by base pairing with the target transcript. However, as the *agr* locus has both positive and negative regulatory effects the situation would appear to be complex.

No regulatory system involving a binding protein has been proposed for control of SpA synthesis. As it is unusual to have regulatory sequences 3 to a structural gene it is unlikely that these "CBS" homologies are associated with a gene upstream of SpA. It was therefore decided to examine the possible role that such an operator sequence could

play in the natural host. The mechanism of catabolite repression in Gram-positive bacteria is largely unknown. Studies on the lactose operon of *S. aureus* have shown that the *lac* genes are subject to catabolite repression (McClutchy and Rosenblum, 1963; Oskouian and Stewart, 1990). However, as cyclic AMP has not beem found in physiologically significant concentrations in *S. aureus* (Blumenthal, 1972; Oskouian and Stewart, 1990) it is unlikely to be involved in this type of regulation. Oskouian and Stewart (1990), provide evidence to support a model for catabolite repression of the *S. aureus lac* operon which involves a negative acting transcriptional regulator which binds to the promoter region of the operon and prevents transcription.

Examination of these gene systems and other catabolite repressed *Bacillus* genes by Weickert and Chambliss, 1990 have shown homologous sequences to exist at the promoter region, which suggested a common regulatory site may be involved in catabolite repression in *B. subtilis*. The consensus sequence of this regulatory site in *B. subtilis* catabolite repressed genes can be seen to be very similar to the CRP site and other known operator sites in *E. coli*.

B. subtilis T G t/a . A . C G . T . t/a C A

E. coli a a . T G T G A . . t a . . t C a m A t w t . w g .

Comparison of the *spa* promoter region with the Gram-positive operator region showed no significant homologies, however the regions upstream of the *spa* promoter designated CBS1 and CBS2 did show homologies, in particular CBS1 (93%). As SpA is not subject to catabolite repression in *S. aureus*, the absence of significant homology in the promoter region was not unexpected. It is, however, possible that the CBS1 sequence upstream of the *spa* gene act as an operator site for another regulatory protein in *S. aureus*.

3.3.3 Signal Sequence

Bacterial proteins can be subdivided in two groups, those that remain in the cytoplasm after synthesis and those exported across the cytoplasmic membrane to locations within the cell envelope, or, external to the cell. This latter group, which includes SpA, are referred to as exported or secreted proteins. All secretory proteins have been found to possess precursor and mature forms. The precursor consists of the mature form with an N-terminal extension of 15-40 a.a's, known as the signal sequence. Several basic models describing the mechanism of protein export have been proposed (Oliver, 1985; Perlman and Halvorson, 1983), which differ slightly in their interpretation of the role of the signal peptide in this process. It has been established that signal sequences are required for protein export and are removed during, or shortly after this process by the precursor processing enzymes, signal peptidase I (encoded by the gene *lep*B), or signal peptidase II (encoded by the gene *lep*A). These proteins are 36 kDa and 18 kDa in size, respectively, and have been shown to be integral inner membrane proteins that span the membrane several times, and possess a large periplasmic domain (Wolfe, *et al.*, 1983; Innis *et al.*, 1984; Moore *et al.*, 1987).

Genetic studies on *E. coli* have identified a number of genes that code for essential components of the export machinery and interact with the signal sequence. Six secretion genes (*sec*) have been identified so far: *sec*A (*prl*D), *sec*B, *sec*D, *sec*E (*prl*G), *sec*F and *sec*Y (*prl*A). Their detection and the evidence for their role in secretion has been reviewed by Schatz and Beckwith (1990) and a model for protein export proposed. This states that during, or soon after, the emergence of the preprotein from the ribosome, secB (acting as a molecular chaperone) will bind to the mature part of the protein, in order to stabilise the export component conformation until the protein associates with later components of the secretion apparatus. The signal sequence appears to be responsible for retarding the folding of the native protein to allow secB to bind. Some proteins appear not to require a molecular chaperone at all, or may use an alternative chaperone protein.

The signal sequence of the precursor peptide-secB complex then interacts with secA and other components in the membrane (this membrane compound includes the inner membrane spanning protein secY and secE) stimulating secA's ATPase activity and the initiation of the translocation reaction. The secD and secF proteins are both integral inner-membrane proteins containing several potential membrane-spanning sequences. They have a large periplasmic domain and mutational studies suggest their involvement in the later stages of translocation.

Signal sequences from both eucaryotic and procaryotic sources have been shown to work in either host system (Oliver, 1985), and bacterial processing enzymes to correctly cleave the signal peptide from a eukaryotic precursor to generate the mature protein (Watts *et al.*, 1983). Thus one might expect to see a significant level of conservation, both structurally and functionally, between signal sequences from diverse sources. However, examination of 277 signal sequences compiled by Watson (1984) reveals no significant homology in either their nucleotide or a.a sequences; with the exception of the signal peptide processing, where Perlman and Halvorson (1983) reported a marked preference for the sequence Ala-X-Ala immediately preceding the cleavage position. The overall length of the signal sequence was also found to vary considerably, from 13 a.a.'s in mouse β-crystallin to 36 a.a.'s in *S. aureus* protein A. However, signal sequences have been found to be homologous in that all consist of three structurally dissimilar regions; (i) a positively charged N-terminal region; (ii) a central hydrophobic region, and (iii) a more polar C-terminal region. Examination of these regions have enabled von Heijne (1983, 1984a, b, 1985) to suggest criteria for a minimal signal sequence.

The signal sequence of the *spa* gene is found to conform to the outer limits of these criteria in most respects. However, at 36 a.a's it is unusually long, in particular the 10 a.a C-terminal region where it exceeds the maximum "limit" by 3 a.a's. The N-terminal region, 11 a.a's in length, carries a net charge of +5 (Figure 3.6) compared with the mean value of +1.7 for prokaryotes (ignoring the α-NH_2 at the N-terminus).

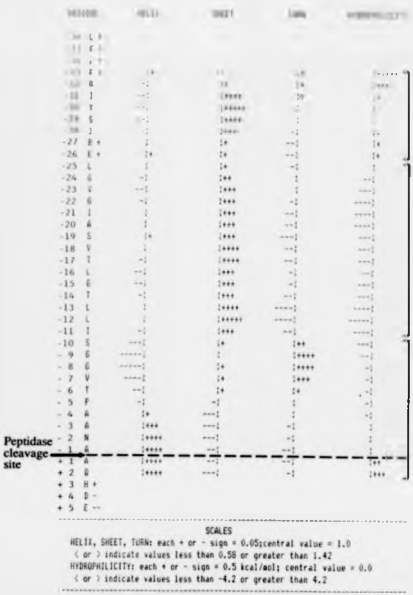


Figure 3.6 Semi Graphic Printout for SpA Signal Peptide

Utilising Chou-Fasman (1978) conformational parameters (turn function PTI) and Kyte-Doolittle (1982) hydrophilicity values.

This charge is thought to participate in binding and orientating the leader sequence with respect to the cytoplasmic membrane surface. The central hydrophobic core (60% hydrophobic residues) consisting of 15 a.a's, is one of the longest prokaryotic core sequences so far elucidated. It has been proposed to function by partitioning the membrane lipids and spanning the bilayer and perhaps act as a target for the signal recognition particle in establishing the translational block. Using Chou-Fasman conformational parameters (1978) it is predicted to be a β-sheet rather than an α-helix (Figure 3.6). Oliver (1985), however, questions the advisability of using rules based on protein folding within an aqueous environment on proteins in the nonpolar environment of the membrane. Von Heijne (1985) predicts that the secondary structure of the hydrophobic core will depend on its length i.e., 8 residues in a fully extended chain structure will have a length of 27Å and therefore be able to span the 25 to 30Å thick membrane interior, whereas it would take a 20 residue helix to span this distance. Therefore, as a helical conformation should be thermodynamically preferred in a non-polar environment, part of the 15 amino acid sequence of the hydrophobic core could be expected to be present in a helical conformation.

The C-terminal region of the signal sequence and the five N-terminal amino-acids of mature protein, are thought to be responsible for cleavage specificity. This C-terminal region consists of a β-turn following the hydrophobic core (thought to be important for peptidase access to the cleavage site; Perlman and Halvorson, 1983), and a signal peptidase recognition site. The *spa* signal sequence C-terminal region has a 5 residue β-turn immediately following the hydrophobic core (Figure 3.6) and, although the C-terminal region is unusually long, it conforms to most of the following preferences observed by Perlman and Halvorson (1983) and von Heijne (1983, 1984a): small neutral residues in positions -1 and -3 (counting from the cleavage site between positions -1 and +1) with *Ala* the most preferred residue at position -1 and common at position +1, but rare in positions -2 and +2 to +5; large polar and aromatic residues are common at position -2; charged and large polar residues are enriched in positions +1 to

+5; Pro is absent from positions ns -3 to +1 but common at position -5 and; *Gly* is found predominately between residues -4 to +1. The SpA signal sequence cleavage region has the residue *Ala* at positions -4, -3, -1 and +1; the large polar residue *Gln* at position +2 and three large charged residues at positions +3, +4 and +5; the large polar residue Asn at position -2, and Pro at position -5, but *Gly* residues are found at positions -8 and -9 rather than between positions -4 to +1 (Figure 3.6).

3.3.4 IgG binding regions

The IgG binding regions of SpA have been defined by Sjodahl (1976, 1977 a and b), as the functional IgG binding units produced upon cleavage of the mature protein with trypsin. These four units: D, A, B and C are consecutively arranged from the N-terminal part of the protein (Figure 3.2).

3.3.4.1 Sequence analysis

Analysis of the predicted SpA a.a. sequence reveals four homologous regions corresponding to the above IgG binding units and an additional region (E), which also exhibits extensive homology with regions D, A, B and C (Figure 3.7). Region E has subsequently been shown to possess weak IgG binding properties (Moks *et al.*, 1986). The five IgG binding regions vary slightly in length, with A, B and C consisting of 58 amino-acids, D consisting of 61 amino-acids and E of 56 amino-acids. Alignment of the IgG binding regions to give maximum homology (Figure 3.7) allows the comparison of both the nucleotide sequence and predicted amino acid sequence of each of the IgG binding regions with one another. This indicates the existence of an homology gradient, where the closer the IgG binding regions are to one another, the greater the degree of homology seen (Table 3.4). The level of homology between regions E and D would at first appear to be an exception to this observation, with region E having greater homology to regions A, B, and C than to region D. However, this anomaly is a result of region D containing three more amino acids than regions A, B,

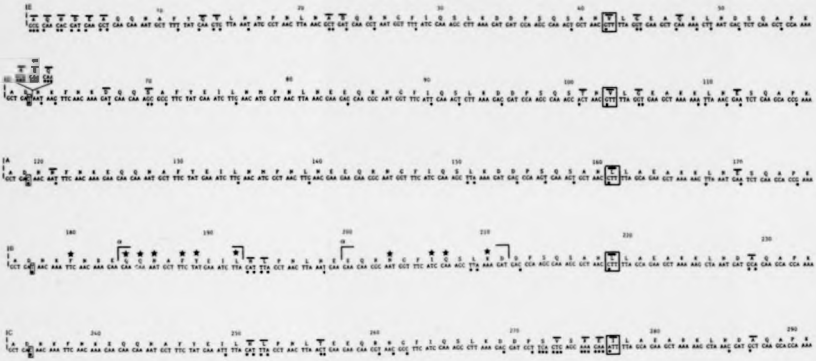


Figure 3.7 Alignment of 5 IgG-Binding Regions of Cowan I NCTC8530

A consensus sequence (not shown) was derived by the alignment of the 5 IgG-binding regions to determine the most commonly occurring a.a or nucleotide at any given point in the alignment. Variations from this consensus are shown on the figure as indicated.

— a.a change from consensus
• nucleotide change from consensus
[] no clear a.a or nucleotide consensus

[ ] denotes residues forming the α helices of the IgG binding domain.

• SpA residues essential for binding to the Fc domain to occur (Deisenhofer, 1981).

Table 3.4 Comparison of the 5 IgG-binding Regions to Show the Existence of an Homology Gradient.

A.

	E	D	A	B	C	TOTAL
E	0 (0)	19 (56)	16 (41)	18 (40)	22 (57)	75 (194)
D		0 (0)	9 (36)	14 (40)	17 (49)	59 (181)
A			0 (0)	6 (15)	11 (38)	42 (130)
B				0 (0)	5 (25)	43 (120)
C					0 (0)	55 (169)
TOTAL						274 794
% HOMOLOGY						81 82

B.

	E	D	A	B	C	TOTAL
E	0 (0)	19 (53)	16 (43)	18 (44)	22 (54)	75 (194)
D		0 (0)	9 (32)	12 (37)	15 (47)	55 (169)
A			0 (0)	5 (16)	10 (36)	40 (128)
B				0 (0)	5 (26)	40 (123)
C					0 (0)	52 (163)
TOTAL						262 777
% HOMOLOGY						80 82

A. *S. aureus* Cowan 1 (NCTC8530)

B. *S. aureus* 8325-4

Total number of matching amino acid residues and nucleotides (figures in brackets), for each of the 5 IgG-binding regions compared with one another, and level of overall homology across the IgG-binding domains is indicated.

and C and five more than region E, where these extra amino acids (nucleotides) are counted as non-homologous. Region E is the least homologous of the IgG binding regions, as it possesses a completely different, six amino acid, N-terminal region from any of the other IgG binding regions. This difference in the N-terminus is almost certainly due to region E being positioned immediately downstream of the signal sequence, for, as mentioned in section 3.3.3, the first five amino acids of the mature protein sequence appear essential for the correct functioning of the signal sequence. It might therefore, be germain to position the start of region E six amino acids downstream of the N-terminus of the mature protein. In order to maintain a degree of consistency with previously published material of Colbert *et al.*, 1984a&b; Guss *et al.*, 1985; Moks *et al.*, 1986; Uhlen *et al.*, 1984 a and c, this has not been done.

Comparison of the number of changed amino acids with the number of changed nucleotides of the identically positioned residues, when the IgG binding regions are aligned and compared in pairs (Table 3.4), shows that the codons have changed at a faster rate than the amino acids. This is indicative of evolutionary pressure to preserve the particular amino acid sequence.

The homology plot (Figure 3.3) also shows the presence of internal homologies within IgG binding regions D, A, B and C. Region A (Figure 3.7) shows the greatest degree of internal homology with 24 of the 30 nucleotides coding for the ten amino acids from residue Asn^{121} to residue Phe^{130} identical to the nucleotide sequence coding for ten amino acids from residue Asn^{138} to residue Phe^{147}.

3.3.4.2 Secondary structure prediction and analysis

The structural interaction between SpA region B and an IgG Fc fragment has been studied using X-ray crystallography at 2.9-Å resolution by Deisenhofer (1981). These studies indicate that SpA fragment B consists mainly of two antiparallel α-helices (though Langone (1982a) mentions a third region of α-helix), the rest of the fragment

consisting of irregularly folded polypeptide. They reveal extensive contact between the two antiparallel α-helices occurring at residues Gln^{128} to Leu^{136} and residues Glu^{144} to Asp^{155} and the CH2 and CH3 domains of the IgG Fc fragment. The residue numbers above correspond to the Cowan I SpA mature protein sequence residue numbers 184, 192, 199 and 211, respectively,as shown on Figure 3.7. This discrepancy in amino acid numbering results from the fact that, for the secondary structure analysis, the amino acid sequence of Cowan I NCTC 8530 is numbered from the N-terminus of region E ,the start of the mature protein. Deisenhofer numbers from the start of region D, this begins 56 amino acids downstream of region E, which was only identified later upon subsequent nucleotide sequencing of the *spa* gene. All numbering of amino acids within this section, (3.3.4), will be according to the NCTC 8530 SpA mature protein sequence displayed in Figure 3.8. Of the 21 amino acids forming the two α-helices the following 10 have been proposed as essential for binding to the IgG Fc domain. These comprise Phe^{180}, four amino acids residues prior to the first α-helix; the first α-helix amino acids Gln^{184}, Gln^{185}, Asn^{186}, Phe^{188}, Tyr^{189}, Leu^{192}; and the second α-helix amino acids: Asn^{203}, Ile^{206}, Gln^{207}; (Figures 3.7 and 3.8)

Comparison of the modified computer prediction of secondary structure for NCTC 8530 SpA fragment B (Figure. 3.8) with the model produced by Deisenhofer (1981) reveals some differences in the extent and position of the α-helix regions in fragment B. The first region of α-helix is three residues longer, Lys^{182} to His^{193}, than Deisenhofer's interpretation, Gln^{184} to Leu^{192}. The second region of α-helix is far shorter than in Deisenhofer's model, consisting of only 6 residues, Glu^{199} to Gly^{204} rather than 12 residues from Glu^{200} to Asp^{211}. The residues from Phe^{205} to Asp^{211} (including two residues considered essential for IgG binding) are predicted to form a B-turn rather than α-helix. A third α-helix region of 15 residues is predicted to occur from residue Leu^{199} to the C-terminal residue, Lys^{203}, of fragment B. Although this third helical region is mentioned by Langone (1982a), Deisenhofer (1981) records this region as having no regular structure. This discrepancy between the computer

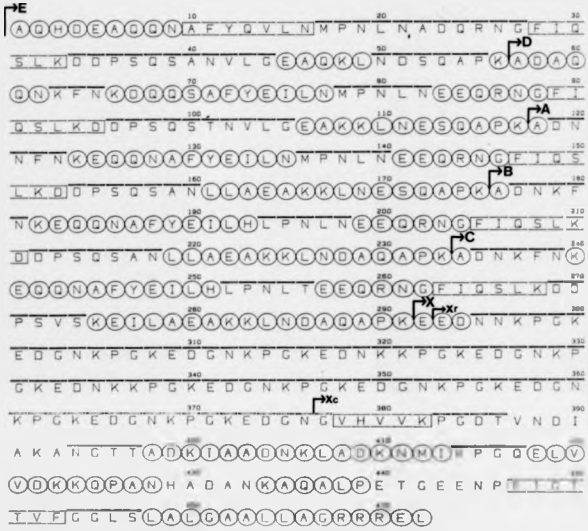


Figure 3.8 **A Secondary Structure Prediction for *S. aureus* Cowan I Protein A**

Circled residues represent α-helix, boxed residues represent β-sheet and residues below a solid bar are proposed to form β-turns. The secondary structure was predicted using the hydropathic profile of SpA (Figure 3.9) and the predicted parameters of Chou and Fasman (1978).

predicted secondary structure for SpA fragment B and that indicated using X-ray crystallography may have been caused by any one, or a combination, of the following factors: Firstly, Deisenhofer obtained 2.8Å resolution data in refining his model for the complex between the Fc fragment and SpA fragment B, which is greater than the peptide bond length of 1.9Å, thus he may not have detected all the existing backbones; this could alter the final structural model. Secondly, Deisenhofer examined the two fragments (and not intact proteins) in their interactive form, whereas the computer prediction is of the full mature SpA molecule in its quiescent state; and the configuration of the protein may alter considerably between these two states. Thirdly, the accuracy of a computer prediction for protein secondary structure depends on the algorithm used and even when accurate with known proteins is of a limited value in predicting "unknown" proteins.

Examination of the hydropathic profile of the two anti-parallel α-helices of region B reveals that both consist of a strongly hydrophilic region followed by a hydrophobic region (Figure 3.8). The first α-helix, consisting of residues Lys^{182} to His^{193} follows a hydrophilic B-turn region from residue Asp^{177}. The first 5 residues of the α-helix are also hydrophilic. The second α-helix (residues Glu^{199} to Gly^{204}) is entirely hydrophilic. It is preceded by a 4 residue hydrophilic β-turn and followed by a hydrophobic 4 residue β-turn , the latter β-turn contains the IgG binding residues Ile^{206} and Gln^{207}. The hydrophilic regions are likely to exist as surface loops in the main, therefore exposing the IgG binding residues within them, and the immediately adjacent hydrophobic region, on the aqueous surface of the protein. The rest of the first α-helix and the β-turn immediately following the second α-helix is hydrophobic, this agrees with Deisenhofers statement that the actual contact between the IgG Fc fragment and SpA is predominantly hydrophobic.

Comparison between the predicted secondary structure of region B and the other four IgG binding regions (Figures 3.7 and 3.8) shows that regions A and C are identical to B, except that the third helical region in region C is one residue longer. Region D

varies in having a 6 residue region of α-helix at its N-terminus, and, in having a shorter region of α-helix at its C-terminal end. It is also unusual in that one of 11 residues considered necessary for IgG binding (corresponding to Asn^{186} in region B), is replaced by a Ser residue. However, this alteration does not effect the helical structure of this region. Region E shows the greatest variation from region B. The principle differences are; (i) it contains only one region of α-helix composed of 9 N-terminal residues positioned at the N-terminus which incorporates only three of the essential 11 IgG binding amino acids (Gln^7, $Gln^8$and Asn^9); (ii) of the 17 residues normally found in helical form in the other 4 IgG binding regions residues 13, 14, 22 and 23 are different; (iii) of the 11 essential IgG binding residues the first residue corresponding to Phe^{180} is altered; (iv) additionally, the hydrophilicity plot (Figure 3.9) of region E (Gln^7 to Gln^{30}) is almost identical to the hydrophilicity plots of the corresponding residues in the other four IgG binding regions. Together with the fact that region E exhibits the weakest IgG binding activity (Moks *et al.*, 1986) these differences suggest that either the altered residues, or the α-helical structure, or both, are crucial to the binding properties of the region.

3.3.5 Cell-Membrane and Cell-Wall Binding Domain

Protein A produced by NCTC8530 is normally bound to the cell wall. The C-terminal domain of SpA, region X, is thought to act as the cell wall anchorage domain for the protein. The position of the N-terminal of region X is defined as the fifth trypsin cleavage site within the protein, and is only accessible in cell wall bound molecules which have been released by lysostaphin digestion. The X region is 181 codons in length and consists of two structurally different domains: Xr- a highly repetitive, hydrophilic, region consisting of a 24 nucleotide unit repeated 10.5 times and; Xc- a non-repetitive sequence, 98 codons in length, concluding with the termination codon TAA. The boundary between the Xr and the Xc region is defined by the end of a 24 nucleotide repeat pattern at nucleotide position 1933 (Figure. 3.10). The Xc region is mainly hydrophilic in nature, except for the C-terminus which consists of a stretch of

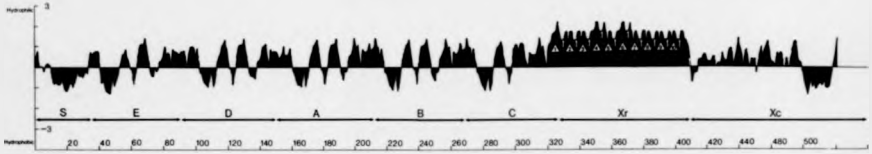


Figure 3.9 Hydropathic Plot of the Amino Acid Sequence of Cowan I protein A

This plot was made using the programs of Larson and Messing (1983) on an Apple II microcomputer, hexapeptide averages were determined from the values of Hopp and Wood (1981). Peaks above the abscissa represent hydrophilic regions, and below, hydrophobic regions. The position of the SpA functional domains are marked.

X1	Glu	Asp	Asn	Asn	Lys	Pro	Gly	Lys
	GAA	GAC	AAC	AAC	AAG	CCT	GGT	AAA
X2	Glu	Asp	Gly	Asn	Lys	Pro	Gly	Lys
	GAA	GAC	GGC	AAC	AAA	CCT	GGT	AAA
X3	Glu	Asp	Gly	Asn	Lys	Pro	Gly	Lys
	GAA	GAC	GGC	AAC	AAA	CCT	GGT	AAA
X4	Glu	Asp	Asn	Lys	Lys	Pro	Gly	Lys
	GAA	GAC	AAC	AAA	AAA	CCT	GGC	AAA
X5	Glu	Asp	Gly	Asn	Lys	Pro	Gly	Lys
	GAA	GAC	GGC	AAC	AAA	CCT	GGT	AAA
X6	Glu	Asp	Asn	Lys	Lys	Pro	Gly	Lys
	GAA	GAC	AAC	AAA	AAA	CCT	GGC	AAA
X7	Glu	Asp	Gly	Asn	Lys	Pro	Gly	Lys
	GAA	GAT	GGC	AAC	AAA	CCT	GGT	AAA
X8	Glu	Asp	Gly	Asn	Lys	Pro	Gly	Lys
	GAA	GAC	GGC	AAC	AAG	CCT	GGT	AAA
X9	Glu	Asp	Gly	Asn	Lys	Pro	Gly	Lys
	GAA	GAT	GGC	AAC	AAG	CCT	GGT	AAA
X10	Glu	Asp	Gly	Asn	Lys	Pro	Gly	Lys
	GAA	GAT	GGC	AAC	AAG	CCT	GGT	AAA
X11	Glu	Asp	Gly	Asn				
	GAA	GAC	GGC	AAC				

Figure 3.10 **Alignment to Show the Repetitive Nature of the Xr Region of NCTC8530 *spa***

20 strongly hydrophobic amino acids followed by a 5 a.a charged tail (Figure 3.9).

This hydrophobic C-terminal structure suggests that this region of SpA may well be associated with the cell membrane (Guss *et al.*, 1984a and b). Similar regions in other Gram-positive cocci surface proteins have been identified and used to predict the membrane anchors (Fischetti *et al.*, 1990). Von Heijne (1981), has suggested that the basic residues present at the C-terminus of a membrane-penetrating protein are designed to help anchor the protein to the negatively charged cytoplasmic surface of the cell membrane. The 5 charged C-terminal residues of SpA may act as a stop-transfer sequence (Blobel, 1980), these sequences are proposed to interrupt the translocation process by preventing a distinct segment of the nascent polypeptide from translocation across the membrane and in so doing will cause the integration of the protein into the membrane.

The 252 nucleotides that constitute the Xr region show a high level of internal homology. Hydropathic analysis of this region reveals that it is extremely hydrophilic (Figure 3.9) and secondary structure prediction suggests the entire region consists of a series of β-turns (Figure 3.8). Alignment of the nucleotide sequence to give maximum internal homology (Figure 3.10) reveals the presence of a 24 nucleotide (octapeptide) unit, repeated 10.5 times. The homologous nature of the region is particularly apparent at the amino acid level, with the 84 codons of region Xr coding for only 6 different types of amino acid: Glu, Asp, Asn, Lys, Pro and Gly. Comparison of the octapeptide repeats revealed that, apart from the presence of an Asn instead of a Gly at residue position three (repeat X1), 6 of the octapeptide residues are identical throughout the Xr region. The two remaining residues, positions 3 and 4, follow a regular alternating pattern with Asn replacing Gly as residue 3, and Lys replacing Asn as residue 4, in octapeptide units 4 and 6 (Figure 3.10 and 3.13). Examination of the sequence at the nucleotide level reveals signs of a 48-nucleotide repeat as well as the 24-nucleotide repeat. There also appears to be evidence of an homology gradient throughout the region with both the N-terminal (X1) and C-terminal (X8, X9, X10 and X11) repeat

units showing greater divergence than the central repeat units. The significance of the Xr region in terms of providing evidence for the stepwise duplication of nucleotide sequences in gene evolution will be discussed further in section (3.4).

Sjoquist (1972b) and Sjodahl (1977a) have reported that SpA is bound to the peptidoglycan of the bacterial cell wall by the X region of the protein. However, although these studies indicate a strong association between the presence of protein A and the presence of peptidoglycan in the experiments undertaken, there is no direct evidence to indicate that the protein A is covalently linked to the peptidoglycan. Similar work undertaken by Pancholi and Fischetti, 1988 on the streptococcal M6 protein, which has a strong C-terminal homology to SpA indicates that this protein is probably intercalated within the cross-linked peptidoglycan and not covalently linked to it, thus casting doubt on the type of association between SpA and the cell wall peptidoglycan.

Uhlen *et al.*, (1984b) has confirmed the role of the X region of the SpA in binding protein A to the cell-wall, in experiments where the *spa* coding for a cell-wall bound SpA was cleaved at the C-terminus of the IgG binding domain C. The resulting truncated gene, lacking the entire region X, when expressed in the same *S. aureus* host system as the full length gene, was found to produce a lower molecular weight, extracellular SpA.

Results presented in Chapter 4 of this thesis for pPA34, a plasmid containing a truncated *spa* gene lacking only the Xc portion of the X region, show that when produced in *E.coli*, the truncated protein was not extracellular. This suggests that either the Xr portion of the X region plays a substantial role in anchoring the protein to the cell or that its presence prevents secretion through the cell wall of a Gram-negative organism.

3.4 Comparison of the Nucleotide Sequences from Three Different *spa* Genes

During the course of this study the complete nucleotide sequence for the *spa* genes from *S. aureus* 8325-4 and *S. aureus* Cowan I (ATCC12598) were determined by Uhlen *et al.*, (1984a) and Colbert *et al.* (1984a and b), respectively.

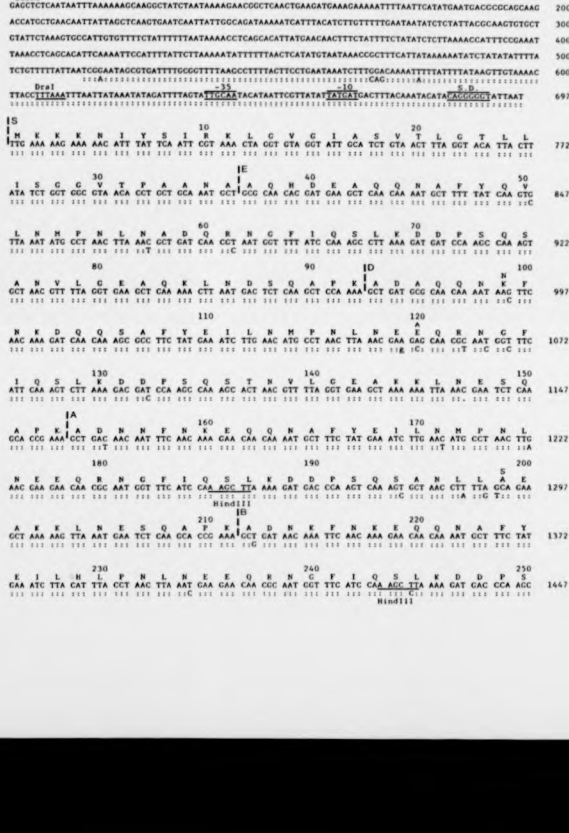
Comparison of the above two nucleotide sequences with the nucleotide sequence of the *spa* gene of *S. aureus* NCTC8530 reported in this thesis, reveal some interesting differences (Figure 3.11), particularly between ATCC12598 and NCTC8530 which are nominally the same strain.

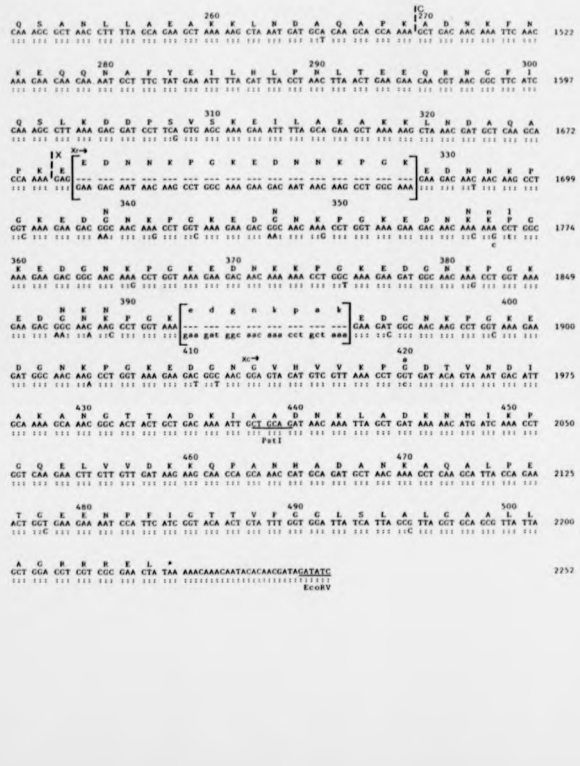
All three *spa* sequences are identical in the region coding for the signal sequence. The nucleotide sequence of the promoter regions of the *spa* gene of 8325-4 and NCTC8530 are also identical. Upstream of the promoter sequence Uhlen *et al.*, identifies a possible hairpin loop transcription termination sequence running from nucleotide position 547 to position 578 (Figure 3.5). However, it is unlikely that this sequence functions as a terminator as not only is the Δ G value of +2.6 kcal mol^{-1} for the proposed stem-loop structure of the RNA transcript not indicative of structural stability, but the comparable region in NCTC8530 differs at 4 of the 18 nucleotide positions forming the stem-loop structure and so will obviate any possible relevant secondary structure. Colbert *et al.* (1984b) have not published nucleotide sequence data for ATCC12598 DNA upstream of the *spa* structural gene, so no comparison is possible. This region has also been identified by Abrahmsen *et al.*, (1986) as a possible stress induced promoter, Figure 3.5.

Comparison of the IgG binding regions of the three genes reveals 20 nucleotide differences between the 8325-4 and NCTC8530 (Figure 3.11), but only a single difference between the two Cowan I *spa* genes, at position 1054. Of these 21 nucleotide differences only 3 result in amino acid changes (at residues 99, 120, and

Figure 3.11 Nucleotide and Amino Acid Sequences of the *spa* Gene

The main sequence printed is the sequence determined in this thesis, of the *spa* from *S. aureus* Cowan I (NCTC8530). It is compared with the sequence of *spa* from *S. aureus* 8325-4 (Uhlén *et al.*, 1984; nucleotide and amino acid differences in upper-case letters) and with the *spa* sequence from *S. aureus* ATCC12598 (Colbert and Anilionis, 1984; nucleotide and amino acid differences in lower-case letters). Colons specify identities; dashes signify deletions. Bracketed regions represent the position of octapeptide repeat units not present in the *spa* gene of NCTC8530 but present in the other two *spa* genes, after alignment to give the least nucleotide mismatches throughout the Xr region. Predicted positions for the -35 and -10 promoters and S.D. sequence are shown, as are the starts of the SpA functional domains.




EcoRV

199) indicating a strong selective pressure to conserve the amino acid sequence. Selective pressure also appears to apply to the secondary structure, as only one of the amino acid changes (residue 120) occurs in one of the antiparallel helices (region D) this is not, however, one of the 11 amino acids actively involved in binding to the Fc fragment of the IgG molecule. None of these amino acid changes alter the predicted secondary structure of the region. Analysis of the homology gradient between the IgG binding regions of 8325-4 compared with a similar analysis on NCTC8530 (Table 3.4 a and b) indicates that a slightly higher degree of homology exists within the 5 IgG binding domains for the latter SpA, at both the nucleotide and amino acid level.

The most significant differences between the three *spa* nucleotide sequences occur in the Xr region of the gene. In NCTC8530 this region consists of a 24-nucleotide unit repeated ten times. The comparable regions in ATCC12598 and 8325-4 contain eleven and twelve 24-nucleotide repeat units respectively. The differing number of 24-nucleotide repeat units makes a direct comparison of the Xr regions difficult, particularly between the two Cowan I "strains" and 8325-4, where it is not possible to determine the position of the insertion/deletion of two of the 24-nucleotide units. Alignment of the nucleotide sequences to give the least nucleotide mismatch (Figure 3.11) positions the extra 48 nucleotides of the Xr region from 8325-4 immediately before the first 24- nucleotide repeat unit of the Xr region from NCTC8530, and the extra 24 nucleotides of the Xr region from ATCC12598 between the eighth and ninth 24-nucleotide repeat unit from NCTC8530. This alignment results in 23 nucleotide differences coding for 6 amino acid differences between the Xr region of NCTC5830 and 8325-4, and 2 nucleotide differences coding for 2 amino acid differences between the Xr region of NCTC8530 and ATCC12598.

Comparison of the three Xr regions reveals evidence for the role of stepwise duplication in the evolution of this repetitive region (Figure 3.12). Both Cowan I Xr regions show a very similar pattern, in having three, 48-nucleotide, repeat blocks (each

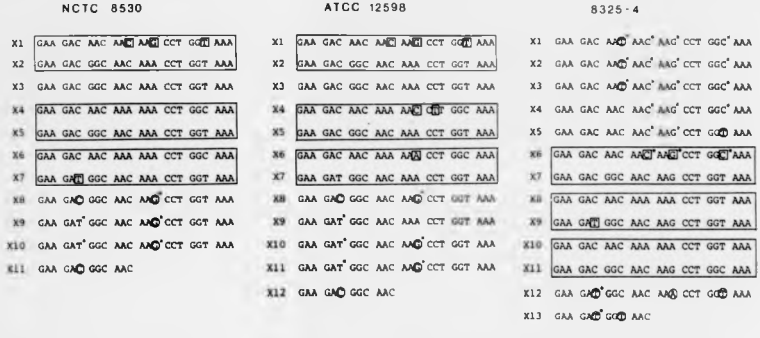


Figure 3.12 Alignment and Comparison of the Nucleotide Sequence of the Xr Regions of the SpA Genes

The octapeptide repeats of the Xr regions have been aligned to give maximum homology. The position of the three 48 mer repeats of each gene are shown (sequence within large rectangular box), and the nucleotide differences between them are boxed (□). The 24 mer repeats proposed to have evolved by duplication of either the 5 or 3 octapeptide of the 48 mer sequence are shown and the mutational divergence from the "parent" octapeptide indicated by a circle (○). The overall evolutionary trends within the Xr region are indicated by an asterisk (*).

interrupted by a 24-nucleotide repeat (X3, which appears to be a direct duplication of X2, the second half of the first 48-nucleotide block) between the first and second 48-nucleotide repeat unit. Repeat unit X9 of ATCC12598 is not present in NCTC8530 and appears to have arisen from a direct duplication of X7. X10 (NCTC8530), X10 and X11 (ATCC12598) appear to have arisen by direct duplication of X9 (NCTC8530). A major dissimilarity occurs between the two Cowan I *spa* genes in repeat unit X4 (residues 5 and 6) where the nucleotide sequence for ATCC12598 is reported as AAC (Asn) and CTT (Leu). Whereas at residues 5 and 6 of the other octapeptide repeats in all three spa genes, the sequence is AAG or AAA (both coding for Lys) and CCT (Pro), respectively. Both of these amino acid changes result from a single nucleotide substitution, creating a unique octapeptide unit in an otherwise highly conserved region. This a.a. substitution dramatically alters the hydropathic plot of the Xr region (Figure 3.13). It is possible that these substitutions could be the result of an error in the sequence data. Alignment of the repeat units of the Xr region from 8325-4 reveals a quite different pattern (Figure 3.12). The three 48-nucleotide repeat units (almost identical to the 48-nucleotide units of both Cowan I genes) are in the C-terminal half of the Xr region. Unlike Cowan I there is only one 24-nucleotide repeat duplication, X12 (which appears most likely to have arisen from a duplication of X9) in the 3′ direction, with the other five 24-nucleotide repeat units X5 to X1 appearing to have arisen by 5′ duplication of X6. Hence the Xr region of Cowan I and 8325-4 appears to have evolved independently or diverged in nucleic acid evolution by a process of stepwise duplication from an original 48-nucleotide unit.

There are two nucleotide differences between NCTC8530 and 8325-4 in the Xc region at positions 2131 and 2182 (Figure 3.11), neither of which alter the amino acid sequence. However, it is of interest to note that the former of these occurs in the highly conserved hexapeptide sequence (PXTGE, as discussed further in Section 3.5). There is a single difference between the Xc region of NCTC8530 and ATCC12598 at position 1956 (Figure 3.11). In the latter sequence, Colbert and Anilionis (1984) report

Most common octapeptide sequence of the Xr region of Cowan I.

	i	ii	iii	iv	v	vi	vii	viii
X2	Glu	Asp	Gly	Asn	Lys*	Pro†	Gly	Lys

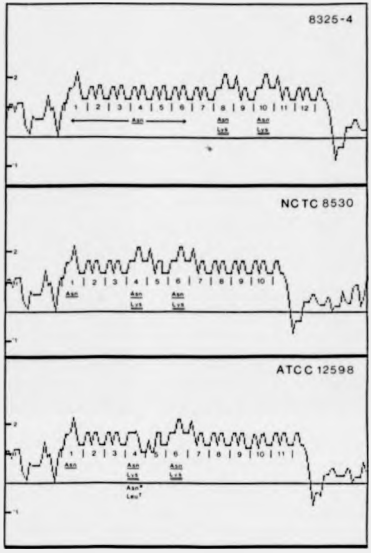


Figure 3.13 Hydropathic Plots to Compare the Peptide Structures of the Xr Region of the Three SpA Molecules

These plots were made using the programs of Larson and Messing (1983) on an Apple II microcomputer. Following the recommendations of Hopp (1986), hexapeptide averages were determined from the values of Hopp and Wood (1981). Peaks above the abscissa represent hydrophilic regions, and those below, hydrophobic regions. The position of the octapeptide repeats are marked and numbered. The most common octapeptide sequence (i.e., X2) is shown. Variations in the octapeptide sequences occur in only 4 a.a. positions and are represented as follows: position iii = ___ , position iv = ___ position v = * and position vi = † , with the a.a. alteration indicated.

a C compared to the G present in NCTC8530, however, this C might be a misprint since they translate the resultant codon (GCT) as Gly (which GGT would give) instead of Ala.

3.5 Homology of SpA with Other Proteins

The IgG binding domains of SpA display some primary sequence homology with the IgG-binding proteins from Group G streptococci (streptococcal protein G, SpG), two cloned and sequenced by Fahnestock (1987); the other partially sequenced by Guss *et al.*, (1986). Although the arrangement and number of IgG-binding regions differs between these SpG proteins, there is complete identity between the proposed FC-binding domains. Despite the fact that SpA and SpG both compete for the same binding region of the Fc receptor there is no significant homology between their primary sequences (Figure 3.14), and a secondary structure prediction of the IgG-binding region C2, did not reveal any α-helices as found in SpA (Guss, 1986). However, there is 80% homology between four Fc-binding regions of SpA (D, A, B and C) and the repeat units A1 and A2 of the SpG sequence reported by Fahnestock *et al.*, (1986b), (Figure 3.14). These units have been proposed to have affinity for human serum albumen (Bjorck, *et al.*, 1987). When analysed for 100% homology, only the pentapeptide LAEAK is common (IgG-binding domains A, B and C of SpA, domain A2 of SpG (Fahnestock *et al.*, 1986b) and domains B1 and B2 of SPG (Guss *et al.*, 1986))

The pentapeptide sequence IQSLK found in all five IgG-binding domains of SpA is also present in the IgG-binding fragment of the murine macrophage-lymphocyte Fc receptor for IgG (amino acid sequence determined by Hibbs *et al.*, 1986). Nucleotide sequence analysis of the gene (Lewis *et al.*, 1986) yields a slightly different pentapeptide sequence ,IQVLK, within the same surrounding peptide sequence determined by Hibbs. Neither IQSLK nor IQVLK is found in either sequence of SpG. This suggests that the binding sites of mamalian macrophages might be similar to those of SpA.

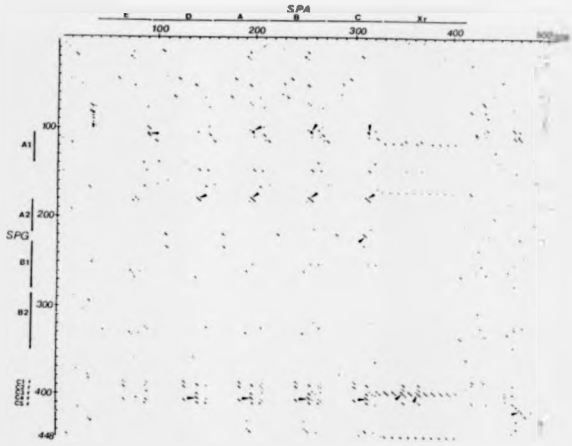


Figure 3.14 Dot Matrix Homology Plot of SpA Against SpG

The primary sequence of SpA (left to right on horizontal axis) is compared with that of SpG (top to bottom). Blocks of 5 a.a. of the SpA sequence were compared sequentially with 5 a.a. of the spG sequence. Where there was identity at a predetermined level (here we show 3 a.a. out of 5, equivalent to 60% homology) a dot was plotted. Regions of homology show as diagonals. The arrow heads (»-) indicate those regions which were present at the 80% homology level. The plots were generated by an Apple II microcomputer using the programs of Pristensky *et al.* (1982). The positions of the Ig-binding domains of SpA (A, B, C, D and E) and SpG (B1 and B2), the cell-wall-binding region, Xr, of SpA, the repeated regions, A1 and A2, and the highly charged repeat structure, C, of SpG are indicated on the axes.

The cell wall attachment domain of SpA shows a repetitive β-turn structure and displays obvious homology with the presumed repetitive wall and membrane attachment sites of SpG. The SpA repetitive β-turn structure however appears to be a structurally sharper turn than the more relaxed repetitive β-turn cell attachment domain of fibronectin (Pierschbacher *et al.*, 1983, Pierschbacher and Ruoslahti, 1984a and b) and more akin to the sharper β-turn membrane anchor sequence of type-M6 protein of group A *Streptococcus* (Hollingshead *et al.*, 1986), and the glycosaminoglycan-binding domain of vitronectin (Suzuki *et al.*, 1985), all of which show marked structural homology with both SpA and SpG in this domain. There is a high extent of sequence homology between fibronectin and the repetitive cell wall attachment domain of SpA. The sequence GNKPG, which occurs seven times in the ten octapeptide repeats forming the cell wall-binding region of SpA, has 80% homology with the sequence GLKPG found in fibronectin. Homology of the primary sequence of the cell wall domain attachment site of SpA with the consensus octapeptide sequence XDGNKPGX, found exclusively in most cell-wall- or membrane-binding proteins, is now evident (Taylor *et al.*, 1986).

Sequence analyses of 11 other surface proteins from Gram-positive cocci has revealed a 16 - 21 a.a. hydrophobic region to be common to the C-terminus (Fischetti *et al.*, 1990). It has also shown a highly conserved hexapeptide sequence (LPXTGE) to be located about nine amino acids N-terminal from the C-terminal hydrophobic region. The SpA protein contains such a sequence (LPETGE), 10 a.a prior to its hydrophobic membrane anchor region. The strong conservation of this hexapeptide sequence at both the DNA and protein level among surface proteins from different Gram-positive cocci suggests that it must be of importance in the anchoring of the surface proteins.

CHAPTER 4

EXPRESSION OF PROTEIN A IN *ESCHERICHIA COLI*

4.1 INTRODUCTION

Successful production of a foreign protein in an organism is dependent on many factors, including; efficiency of gene transcription, efficiency of mRNA translation and correct processing of post-translational polypeptides. Problems at the level of transcription are often caused by the gene promoter not functioning properly in the foreign host, or, even if the promoter functions normally, it may be too weak to direct sufficient transcription to produce a recoverable amount of the protein. Efficient translation of the mRNA relies on the stability of the message, its ability to bind ribosomes and the presence of sufficient quantities of the appropriate tRNA (section 3.3.1). Even if the foreign gene is successfully expressed, the desired protein recovered from the host cell may be altered from that produced by the natural source. The host organism may incorrectly process the polypeptide in several ways. It may lack the enzymes necessary to recognise specific cleavage sites, essential for the development of the mature protein from a precursor form. Alternatively it may possess enzymes that cleave the protein at inappropriate positions and so produce degraded polypeptides. The host cellular transport mechanisms may fail to recognise signals carried by the foreign protein, so preventing its localisation to the appropriate position for processing. A fault in any one of these three major processes will severely limit or prevent production of the desired protein.

The detection of the *spa* gene in a *S. aureus* gene bank was dependent upon I^{125} labelled IgG binding to SpA released from lysed *E. coli* colonies (Duggleby and Jones, 1983). The success of this screening procedure demonstrated that *E. coli* HB101 was capable of producing an active form of SpA.

Preliminary experiments were carried out by Duggleby and Jones (personal communication) using two different plasmids: pBR327 (medium copy no. of 150 per cell) and pUC8 (high copy number of 500 to 600 per cell, and carrying the *lac* promoter

upstream of the multiple insertion site region); to express a 2.012 bp fragment of the *spa* gene. These constructs were designated pPA12 and pPA16 respectively (further details in Section 4.2.1.1). The level of SpA production achieved (quantitated using an ELISA system rather than a modified monorocket technique) was no higher than that attained using *S. aureus* Cowan I. The four-fold increase in SpA yields observed for pPA16 ($2.4\mu g ml^{-1}$) over pPA12 ($0.6\mu g ml^{-1}$) is not consistent with an increase in expression due to the presence of the *lac* promoter. Indeed, the increase could be solely attributable to a gene dosage effect, as a consequence of the higher (3-4 fold) copy number of pPA16 compared to pPA12, and suggested that, transcription from the *lac* promoter had no discernible affect on *spa* expression.

In this Chapter, experiments are described which aimed to improve the level of SpA production in *E. coli*. This involved the construction of various recombinant plasmids containing either authentic or modified (by deletion of structural/regulatory components and/or alteration of the translational initiation codon) *spa* genes. All these manipulations were made possible by the availability of the nucleotide sequence described in Chapter 3.

4.2 RESULTS

4.2.1 Protein A Production in *E. coli*

Aside from investigating the genetic factors which affect *spa* expression, an investigation of the physiological characteristics of high level SpA production was also required. Accordingly, the studies described in this section describe the levels of SpA production obtained with the initial recombinant plasmids, the effect of production on the host and the characteristics and cellular localisation of the IgG-binding polypeptides produced. A more detailed examination of gene expression is described in the later sections, 4.2.2 and 4.2.3.

4.2.1.1 Plasmid Constructions Utilised

Various recombinant plasmids carrying *spa* specific DNA were initially constructed to evaluate the effects of a number of parameters on SpA production levels in *E. coli*. These parameters included plasmid copy number, the presence or absence of C-terminal *spa* codons, and the effect of placing a strong promoter (*lac*) at two different locations 5′ to the *spa* gene. Restriction maps of these pUC8 based plasmids (pPA16, pPA30, pPA31 and pPA34) are illustrated in Figure 4.1 and their pertinent features listed below.

pPA12: A 2.012 kb *Eco*RI/*Pst*I fragment inserted between the appropriate sites of pBR327. This plasmid, therefore has a relatively low copy number (150 copies per cell; Chambers *et al.*, 1988) and, aside from the *spa* promoter, has no recognised strong promoter 5′ to the staphylococcal insert. The *spa* gene itself lacks 69 codons at the 3′ (*Pst*I) end, which are replaced by 100 codons derived from the 3′ end of the vector *bla* gene.

pPA16 and pPA34: Composed of either a 2.012 kb *Eco*RI/*Pst*I fragment (pPA16) or a 1.404 kb *Dra*I/*Pst*I fragment (pPA34) inserted between the *Eco*RI and *Pst*I sites (pPA16), or the *Sma*I and *Pst*I sites (pPA34), of the multicopy plasmid pUC8. Both plasmids have high copy numbers (600 copies per cell; Chambers *et al*, 1988) and lack 69 codons from the 3′ end of their *spa* genes. In both cases fusion has occurred between the remaining 438 *spa* codons and 14 nonsense codons from the vector *lacZ* region (see 3.2.3). The two plasmids differ only in that the translational initiation codon of the *spa* gene carried by pPA16 is separated from the vector *lac* promoter by an additional 608 bp of staphyloccocal DNA compared to pPA34.

pPA30 and *pPA31*: Derived by insertion of either a 2.252 bp *Eco*RI/*Eco*RV fragment (pPA30) or a *Dra*I/*Eco*RV fragment into pUC8 cleaved with *Eco*RI/*Sma*I and *Sma*I, respectively. Plasmids pPA30 and pPA31 are identical to pPA16 and pPA34,

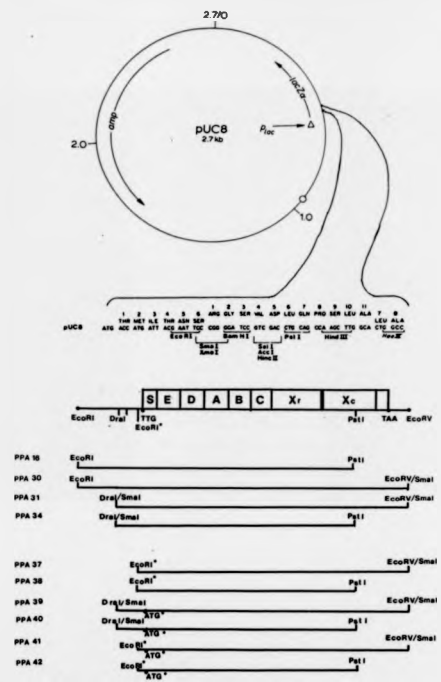


Figure 4.1 Construction of the SpA Plasmids pPA16→pPA42

Mutagenic changes from the original *spa* sequence are indicated by an *

The Restriction Enzymes (RE) used to cleave *spa* and pUC8 prior to recombination are indicated. Where two different RE's were used to produce compatible sites, for example *Dra*I/*Sma*I, the first RE is that used to cut *spa* and the second RE that used to cut pUC8.

respectively, except for the fact that both carry the entire *spa* structural gene, including the *spa* translation termination codon (TAA) and thus do not produce an SpA fusion protein.

All of the pPA plasmid constructions contain no recognisable transcription termination sequences before the 3′ end of the cloned *S. aureus* DNA fragment.

4.2.1.2 Preliminary Estimation of the Level of SpA Production

The four plasmids, pPA16, pPA30, pPA31 and pPA34, were transformed into JM83, to allow constitutive expression of the pUC8 *lac* promoter. The cultures were grown according to the conditions set out in section 2.2.1.1. The SpA was recovered from the cells and the amount determined using monorocket electrophoresis; as detailed in sections 2.2.13.2 and 2.2.14.2 respectively.

To obtain a crude idea of the levels of SpA being produced by the various recombinant clones, each was grown overnight in a 10 ml culture volume. The levels of SpA produced are summarised in Table 4.1. Estimates of the SpA levels produced by the different recombinant clones varied considerably from each experiment, as evidenced from the large standard deviation values obtained. To determine whether there was a significant difference in the levels of SpA production between the different clones, a statistical comparison was necessary. A non-parametric method of analysis (Siegel, 1956) was employed as the SpA results did not comply with one of the main requirements for a parametric test, in that the populations to be compared did not have the same variance. The Kruskal-Wallis one-way analysis of variance was used to determine the probability that a significant difference existed between the populations in question. If a significant difference was indicated, then the Mann-Whitney U test was used to determine the level of significance between the SpA results from each of the plasmids. Analysis of the data used to calculate Table 4.1, showed that: (i) there was no significant difference at the 10% level between SpA levels from Cowan I and pPA16; (ii) that the SpA levels produced by pPA30 compared with those of Cowan I

TABLE 4.1

Comparison of the SpA Levels Achieved by the Different

Plasmid Constructions in 10 ml and 500 ml Culture Volumes

Culture Size	SpA $\mu g\ mg^{-1}$ Cells Dry wt.				
	Cowan I	pPA30	pPA16	pPA31	pPA34
10 ml	29.6	5.5	32.2	65.5	123.7
	(7.6)	(0.76)	(7.8)	(13.2)	(30.2)
500 ml	-	-	59.5	116.0	279.2
	-	-	(11.8)	(22.0)	(29.8)

Results from o/n and timed runs at 37°C, 100 rpm, 10 ml in 20 ml container and 500 ml in 2 L container. SpA was released from JM83 cells using lysozyme Brij/Doc lysis. SpA was released from Cowan I cells using lysostaphin

10 ml cultures: Cowan I and pPA30, samples from 6 sets of experiments.

pPA16, pPA31 and pPA34, samples from 11 sets of experiments.

500 ml cultures: samples from 5 sets of experiments.

Figures in brackets ( ) = std. deviation values (n-1) to show sample spread.

and pPA16 were significantly different at the 5% levels (significant, 1 in 20 chance of error), and; (iii) that the rest of the comparisons between plasmid SpA levels were significant at the 1% level (highly significant, 1 in 100 chance of error).

Comparison of the mean SpA levels obtained for each plasmid derived from pUC8, indicated that deletion of either the upstream *spa* non-coding region (*Eco*RI to *Dra*I, i.e., pPA31) or the 69 codons from the 3′ end of *spa* (*Pst*I to *Eco*Rv, i.e., pPA16), or both (pPA34) significantly increased the levels of SpA attainable. Taking the results obtained with pPA30 as the baseline (i.e., a plasmid carrying the whole *spa* structural gene and upstream 608 bp non-coding region), deletion of 69 codons from the *spa* 3′ end (pPA16) resulted in a 6-fold increase in the levels of SpA produced. Alternatively, deletion of the 608 bp of non-coding DNA from between the vector *lac* promoter and putative *spa* promoter (pPA31) resulted in a 12-fold increase in SpA production. A combination of both deletions (pPA34) resulted in a 24-fold increase in the levels of SpA obtained.

4.2.1.3 Time Course of SpA Production

The previous results represented the level of SpA production achieved following growth of each clone into stationary phase. A Further series of experiments were, therefore, undertaken on cells carrying pPA16, pPA31 and pPA34, in which the level of SpA production at various points in the growth phase was monitored. A 2 L falsk containing 500 ml of prewarmed 2xYT broth, ampicillin (50 $\mu g ml^{-1}$) was inoculated with JM83 [SpA] from an overnight culture to give a starting density of 0.02 mg cells per ml. The cultures were shaken at 160 rpm, 37^{o}C, and 10 ml samples taken at regular intervals, lysed and SpA levels determined as detailed in sections 2.2.13.2 and 2.2.14.2.

Under these growth conditions, an overall increase in SpA production was observed in cells carrying all three plasmids (Table 4.1). A growth curve of the amount of SpA recovered from lysed cells versus number of cells present, was plotted (Figure 4.2).

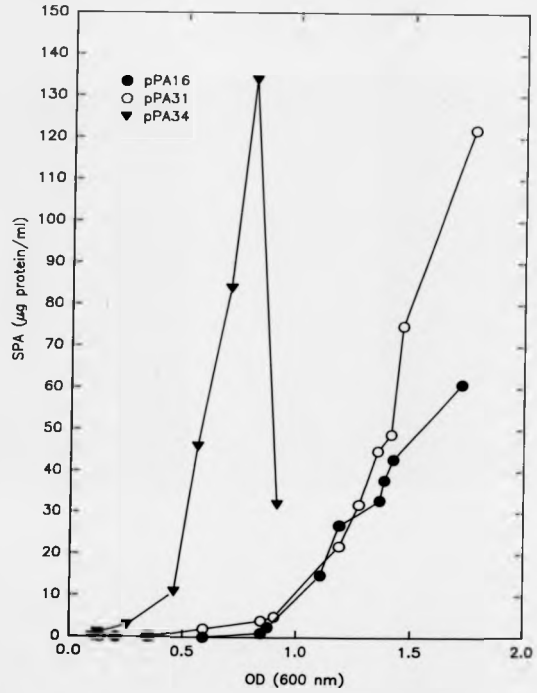


Figure 4.2 Change in SPA levels of 3 recombinant clones as cell density increases

From the graph it can be seen that after an initial lag phase, during which cell growth increased at a far greater rate than SpA production, SpA production increases in proportion to cell growth. The rate of SpA production directed by plasmids pPA16 and pPA31 was virtually identical until a cell density of 0.45 mg ml^{-1} was attained, whereupon, SpA production by cells carrying pPA16 was seen to decrease considerably. The rate of SpA production directed by pPA34 was consistently higher than that seen for the other two plasmids. However, between the penultimate sampling time (6hrs after inoculation), and the last sampling time (overnight, 23.5 hours after inoculation), the levels of SpA present (per mg ml^{-1} cells) dropped from 135.5 µg ml^{-1} to 31.5 µg ml^{-1}. Analysis of the growth medium from the 23.5 hour samples from all three JM83 [pPA] cultures showed no SpA present in that of cells carrying pPA16, SpA levels of 1 µg ml^{-1} in that of cells harbouring pPA31 and 120 µg ml^{-1} in that of cells containing pPA34. Whereas SpA was almost exclusively intracellular in cells containing pPA16 and pPA31, after 23.5 hours incubation under the growth conditions above, almost 80% of the SpA produced by cells carrying pPA34 was found to be extracellular.

4.2.1.4 SpA production and Cell Fragility

Further investigation to determine whether SpA produced by pPA34 containing cells was actively secreted from the cell or was present in the culture medium, solely through cell lysis, were performed. Repeat growth experiments, with conditions identical to the above excepting a reduction in shaking speed, revealed no decrease in cellular SpA levels during growth, and extracellular SpA levels of less than 5 µg ml^{-1} when shaken at speeds less than 120 rpm. As the shaking speed was increased above this level, so the extracellular levels of SpA increased. Determination of the molecular sizes of the IgG binding polypeptides produced showed an identical banding pattern for both intracellular and extracellular SpA. Analysis of the cells under a light microscope showed a high percentage to be lysed "ghosts" (Section 4.2.15). It was concluded that the SpA produced was intracellular in nature, but due to the fragile nature of the host cells, were easily released into the external medium.

The fragile nature of JM83 when carrying an SpA plasmid appeared to correlate directly with the amount of SpA produced. Cell viability was a particular problem with JM83[pPA31] and JM83[pPA34], both during culture growth and during storage of cells (viability problems during large scale cultivation of the latter clone will be discussed further in section 4.2.4). The maximum rpm at which a 500 ml JM83[pPA] culture could be shaken with less than 5% SpA released into the media, when grown under the conditions already stated, was 220 rpm for JM83 [pPA16], 170 rpm for JM83 [pPA31], and 160rpm for JM83 [pPA34]. This contrasts with JM83[pUC8], which could be shaken at 300 rpm with no apparent lysis of the cells. It was found that whereas JM83 [pPA16] remained viable on a 2xYT (ampicillin 50 $\mu g\ ml^{-1}$) plate stored at $4^{o}C$ for approximately 2 months, JM83 [pPA31] and JM83 [pPA34] were found to have lysed after 5 days and 3 days respectively. When incubated at $37^{o}C$ for more than 14 hours both the latter developed a second smaller colony type, which was not resistant to ampicillin and did not produce SpA. No plasmid was detected in these cells. Storage at $-70^{o}C$ revealed a 25%, 30%, and 50% loss in viability for JM83 [pPA16], JM83 [pPA31] and JM83 [pPA34] respectively, compared with a 10% loss in viability in JM83[pUC8] stored for the same length of time.

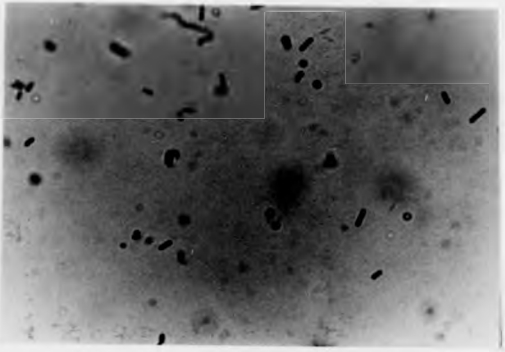
4.2.1.5 SpA Production and Cell Morphology

Examination by light microscopy of a 5 µl sample of SpA producing cells, from a 10 ml culture shaken for 6 hours at $37^{o}C$, magnification 100, revealed a difference in the cell shape for each of the high producing SpA clones (Figure 4.3). JM83[pPA16] appeared almost identical to JM83 [pUC8], a 2-3 µ rod, it was the same length but slightly ovoid in shape. The 5 µl sample of JM83[pPA31] contained 50% normal cells (Figure 4.3), the rest of the cells showed varying degrees of elongation and distortion of the cell wall. Many contained a denser region which is normally indicative of an inclusion body. The 5 µl sample of JM83[pPA34] also contains a mixed population; consisting of 10% normal cells, 90% slightly ovoid cells. The fact that the presence of

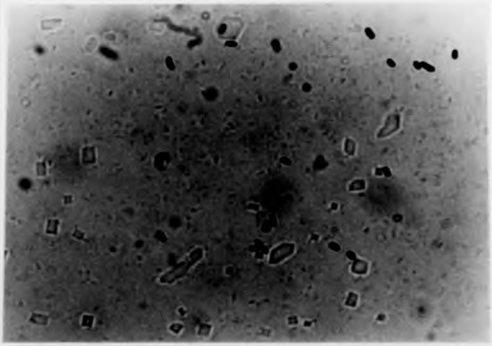
Figure 4.3 **Cell-Morphology of JM83 [pUC8], JM83 [pPA16], JM83 [pPA31] and JM83 [pPA34]**

Gram stain of: A: JM83 [pUC8], B: JM83 [pPA16], C: JM83 [pPA31] (to show the frequency of elongated cells in the culture BG) D: JM83 [pPA31] (to show the extent of elongation of the cells) E: JM83 [pPA34]

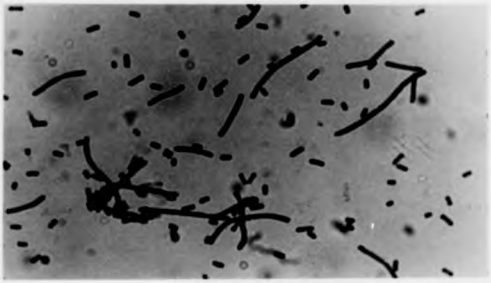
A.



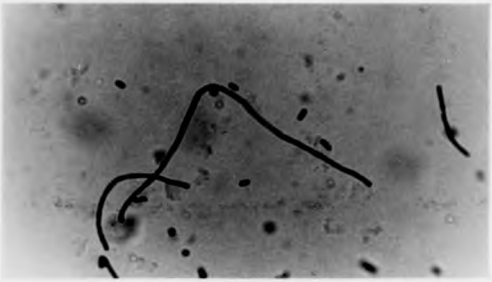
B.



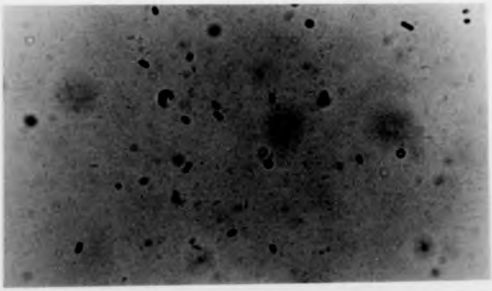
C.



D.



E.



pPA31 caused a far greater distortion of the cell structure than did the presence of pPA34, would seem to indicate that the SpA cell membrane binding region, present in pPA31, might be interacting with the *E. coli* outer membrane. This interaction may also explain the elongation of the cells, as the SpA may be interfering with membrane sites involved in the process of cell division.

That such an interaction occurs is corroborated by experimental evidence obtained from comparing the difference in SpA yields between cells that have been disrupted using a lysozyme/detergent method with those disrupted using sonication (Table 4.2). The former method involves centrifugation to separate the cell debris from the soluble cell fraction, allowing the latter fraction to be collected. Any SpA adhering to the cell membranes, or cell wall, would also be removed. Sonication will fractionate the cell-membranes and cell-wall to such a degree that far less cell debris is pelleted during centrifugation. Cellular disruption, using sonication rather than lysis, results in more than an 80% increase in the recovery of the full-length SpA coded for by pPA31; compared with an increase of only 30-40% in the recovery of the truncated SpA coded for by both pPA16 and pPA34. These results suggest that the proposed membrane binding region of SpA, coded for by pPA31, does associate with the *E. coli* cell-membrane or cell-wall.

4.2.1.6 Production in Different *E. coli* Hosts

The poor viability of the two high expression clones, JM83[pPA31] and JM83[pPA34] would make successful production of SpA in large scale cultures difficult. As *E. coli* strains are known to vary in their susceptibility to cell lysis, the use of a different *E. coli* host might result in greater plasmid and cell stability, without a significant reduction in SpA production. The two high SpA producing plasmids pPA31 and pPA34 were transformed into the *E. coli* strains JM101, MC1061 and GM242. The transformed strains were then cultured in 10 ml volumes in the standard manner, the SpA recovered using the lysozyme method (Section 2.2.13.2) and the levels of SpA compared by monorocket electrophoresis (2.2.14.2). Statistical analysis of the results,

TABLE 4.2

Comparison of the SpA Levels Obtained When JM83 [SpA] Cells are Disrupted Using Either Sonication or Lysozyme/Detergent Lysis

Plasmid	SpA levels $\mu g mg^{-1}$ cells dry wt.		
	Lysozyme/ detergent	Sonication	% increase in SpA obtained using sonication
pPA16	36.2	49.7	37.3
	(0.57)	(0.71)	(3.6)
pPA31	58.6	107.6	83.6
	(12.1)	(20.4)	(5.6)
pPA34	148.8	196.2	31.9
	(22.3)	(27.4)	(1.5)

sample size = 6

std deviation values in brackets (n-1)

summarized in table 4.3, showed that there was no significant difference at the 5% level between the amounts of SpA produced by each of the *E. coli* hosts carrying the same plasmid. The viability of the cells following storage was again found to be directly correlated to the amount of SpA produced; as SpA levels increased, so cell viability decreased.

4.1.2.7 Sizes of IgG binding Polypeptides Produced

The sizes of the IgG-binding polypeptides produced by the different SpA clones were determined by Western blotting. Cell-free extracts were prepared from cultures by lysis with lysozyme and detergent, heated with an SDS containing running dye to destroy secondary structure and fractionated on an SDS- polyacrylamide gel. The proteins were transferred electrophoretically to nitrocellulose filters, which were then washed with 5% BSA to prevent non-specific binding and probed with ^{125}I-labelled human IgG. After exposure to the probe, the filters were washed, dried and autoradiographed. (Details in Methods, sections 2.2.15, 2.2.16 and 2.2.17).

Analysis of the cell-free lysates from HB101[8G4] and JM83[pPA30], both of which contain an SpA fragment capable of coding for the entire gene, showed the presence of more than one IgG-binding polypeptide in both lysates. The number of IgG- binding bands present were found to vary between samples from three (Molecular weight bands 47 kDa, 43 kDa and 14.5 kDa) up to six (tracks 4 and 5, Figure 4.4), though the position of the peptides remained constant relative to one another. The 14.5 kDa IgG-binding band was also shown to be present in the control samples (lysates from identical *E. coli* host [plasmid] system lacking only the SpA inserts). This band was found to be identical in size to the lysozyme band in the protein marker sample. A sample containing buffer and lysozyme, when analysed using the same Western blotting procedure, was found to produce a 14.5 kDa band. Therefore, the 14.5 kDa IgG-binding band in the *E. coli* [SpA] lysates was considered to be lysozyme and not an SpA polypeptide. The smaller molecular weight peptides are thought to be a result of degradation of the mature SpA,

TABLE 4.3

Alternative *E. coli* Hosts

Plasmid	SpA Levels $\mu g mg^{-1}$ Dry Weight *E. coli* Strain			
	JM83	JM101 +inducer (IPTG)	MC1061	GM242
pPA31	65.6	51.4	57.3	72.4
	(13.2)	(8.7)	(6.4)	(15.6)
pPA34	123.7	86.4	86.2	106.6
	(15.1)	(21.6)	(11.8)	(17.2)

Results from 10 ml o/n cultures, 37°C, 200 rpm. 1mM IPTG added to JM101 culture

JM83 and MC1061; 11 samples

JM101 and GM242; 6 samples

std. deviation values in brackets (n-1)

Figure 4.4 **Analysis of IgG Binding Polypeptide Produced by HB101 [8G4] and JM83 [pPA30]**

Lanes contained:

(1) Sigma Protein A, 100 µg/ml,
(2) Sigma Protein A 10 µg/ml
(3) JM83 [pUC8],
(4) JM83 [pPA30],
(5) JM83 pPA30] & pmsf,
(6) Lysozyme
(7) HB101 [pBR327]
(8) HB101 [8G4],
(9) HB101 [8G4] & pmsf

as their number is found to increase in proportion with the length of the cell growth period, (Figure 4.8, reported in greater detail at a later point). Degradation is possibly caused by one of the many proteases known to be located within the *E. coli* periplasm. Degradation was unlikely to be the result of *E. coli* serine protease activity as the inclusion of the serine protease inhibitor, phenylmethylsulphonyl fluoride (PMSF, 0.1mM final concentration), in cell suspensions prior to lysis caused no alteration to the banding pattern observed (Figure 4.6), though a few serine proteases are not sensitive to this reagent.

Analysis of the IgG-binding peptides produced by JM83[pPA31] produced an identical banding pattern to that seen for 8G4 and pPA30. A similar banding pattern was present in lysates from JM83[pPA16] and JM83[pPA34]; the only significant difference being a decrease in the size of the largest molecular weight band from 53 kDa to 48 kDa (Figure 4.5, Table 4.4). This difference in size is approximately 2 kDa less than would be expected for the deletion of the C-terminal region in the latter constructs, and is probably indicative of the limits of accuracy imposed upon the system through trying to cover a wide range of molecular weight polypeptides at inconsistent concentrations. The largest SpA polypeptide band (53 kDa) produced by cells carrying either pPA30 or pPA31 was indistinguishable from the native Cowan I SpA mature peptide, whereas the largest SpA polypeptide from cells containing pPA16 and pPA34 (immunoblot for pPA34 shown only, Figure 4.8) can be seen to have a higher relative mobility. This appears to indicate that the largest molecular weight polypeptide represented the processed (mature) SpA molecule, rather than preprotein form (before cleavage of the signal sequence).

N-terminal amino acid analysis of the IgG-binding polypeptides produced by pPA34, was undertaken by Mr Roy Hartwell, PHLS, CAMR, using an automated Edman degradation technique as detailed in Section 2.2.20 and showed that all the polypeptides possessed the same N-terminal sequence, that of the mature SpA. This information both confirms that the SpA signal sequence cleavage site is recognised and the

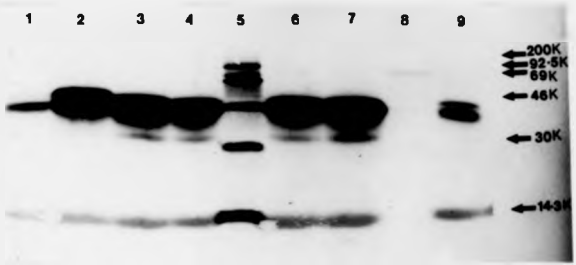


Figure 4.5 **Analysis of IgG-binding Polypeptides Produced by JM83 [pPA16], JM83 [pPA31] and Jm83 [pPA34]**

Lanes contained

(1 & 2)	JM83 [pPA31]	(3 & 4)	JM83 [pPA34]
(5)	Molecular weight marker	(6 & 7)	JM83 [pPA34]
(8)	*S. aureus*, wood 46	(9)	JM83 [pPA16]
(10)	Molecular weight marker		

TABLE 4.4

IgG-Binding Peptides Produced by

the JM83 [SpA] Clones

Molecular weight of IgG-binding peptides present in lysates from the SpA plasmids					
Cowan1	8G4	pPA30	pPA16	pPA31	pPA34
56,000	54,000	54,000	48,000	53,000	48,000
	46,000	45,000	45,000	45,000	45,000
	43,000	43,000	42,000	42,000	43,000
	42,000	42,000	14,000	14,000	32,000
		14,000			28,000
					14,000

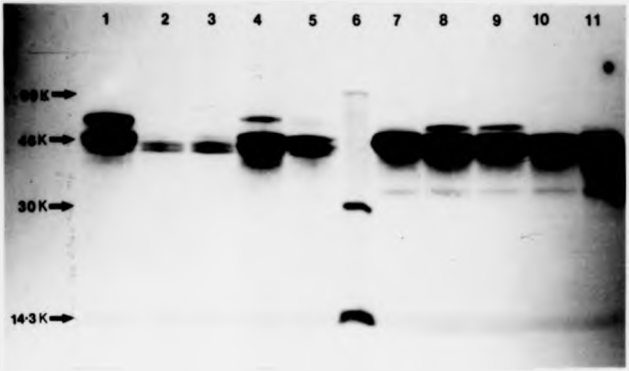


Figure 4.6 **Effect of (i) Lac Control on SpA expression and ii. the Presence of (pmsf) on the Degradation of the IgG-Binding Polypeptides of JM83 [pPA31 and pPA34]**

Lanes contain:

(1) JM83 [pPA31] earlier sample positive control
(2) JM83 [pPA31],
(3) JM83 [pPA31] and pmsf
(4) JM83 [pPA31] and [pNM52]
(5) JM83 [pPA31] and [pNM52] and IPTG
(6) Molecular weight marker,
(7) JM83 [pPA34] earlier sample positive control
(8) JM83 [pPA34]
(9) JM83 [pPA34 and pmsf
(10) JM83 [pPA34] and [pNM42]
(11) JM83 [pPA34] and [pNM52] and IPTG.

preprotein correctly processed by *E. coli* enzymes, and also helps localise the SpA regions where proteolytic cleavage may occur. As the N-terminus has been shown to be intact in the smaller SpA polypeptides, and it is unlikely under the denaturing conditions of the running gel that the formation of a variable secondary structure could account for all the smaller SpA polypeptide bands, any truncated polypeptides are likely to have resulted from proteolytic cleavage occurring in the C-terminal region of the protein. As the banding patterns for the smaller IgG-binding polypeptides are identical for the lysates from both the full length and truncated SpA clones, this infers that proteolytic cleavage is occurring at the N-terminal side of the amino acids corresponding to the *Pst*1 site in the nucleotide sequence. Comparison of the molecular wts obtained for the degraded SpA polypeptides with the amino acids predicted from nucleotide sequencing of the gene, places the C-termini of the 45Kd, 43Kd and 42Kd SpA polypeptides within the first 27 amino acids of the Xc region, and the C-terminus of the 32Kd polypeptide on the boundary between IgG-binding region, C, and the Xr region. This is corroborated by a personal communication from Rene Arenson, Dupont, reporting the presence of 26 lysines in the C-terminal region of 76% of the SpA derived from pPA34, which indicates that the majority of the SpA polypeptides retain the Xr region.

4.2.1.8 Degradation of IgG-binding Polypeptides

Growth experiments on JM83[pPA31] and JM83[pPA34] were carried out to determine the rate of degradation of the IgG-binding polypeptides. One litre of prewarmed (37^{o}C) 2xYT, ampicillin 50 $\mu g\ ml^{-1}$, was inoculated with 7 ml of an overnight bacterial culture to give a starting OD_{450} of 0.1. The cultures were shaken at their optimum SpA production speed, at 37^{o}C, and a 10 ml sample taken at hourly intervals. The SpA was recovered from the cells using the lysozyme/detergent procedure, detailed in methods Section 2.2.13.2, and the amount of SpA determined by monorocket electrophoresis, as detailed in Section 2.2.14.2. The cell-free lysates were then diluted to give a maximum concentration of 100 $\mu g\ ml^{-1}$ and the IgG-binding

polypeptides determined using the Western blotting and radioimmunoassay procedure detailed in sections 2.2.15, 2.2.16 and 2.2.17.

Analysis of the gels, Figure 4.7 and 4.8, reveals that the rates of degradation for two different mature SpA polypeptides, produced by cells carrying pPA31 and pPA34 were different. The lysates from JM38 [pPA31] (Figure 4.7) show both the 53 kDa and 45 kDa polypeptide present in the 1 hour sample; in the 2 hour sample the 53Kd band appeared stronger, however, closer examination of the gel shows evidence for an air-bubble being present during the Western blotting procedure resulting in poor transfer of the 45 kDa band in tracks 4 and 5. The 43 kDa band first appeared in the 3 hour sample, and evidence of further degraded polypeptides was seen in the 7 hour sample. The first IgG-binding polypeptide seen in the lysate from JM83 [pPA34] was the 45 kDa band, which was seen in the sample taken immediately after inoculation of the 1 litre culture. The 48 kDa band, thought to represent the mature form of the truncated SpA, was not visible until the 3 hour sample and was then seen to increase until, in the 6 hour sample, it had a similar intensity to the 45 kDa band (which appeared to have remained at a constant level since the 3 hour sample). This alteration in the proportions of the 48 kDa and 45 kDa bands may have resulted from further degradation of the 45 kDa band, as a 43 kDa band first became apparent in the 4 hour sample, and was seen to increase in intensity, whilst the intensity of the 45 kDa band remained constant. From the 6 hour and later samples, the number of small IgG-binding polypeptides increased dramatically. The 48 kDa band remained at a fairly constant level, taking into account the dilution factor, until the 16 hour sample, when its concentration decreased, along with the 28 kDa band.

During the course of the experiment the levels of SpA were found to decrease from a level of 324 $\mu g\ ml^{-1}$ in the 14 hour sample to a level of 246 $\mu g\ ml^{-1}$ in the 16 hour sample. This may be due to lysis of some of the cells , or alternatively, to continued degradation of SpA already present in the cells without the further synthesis of new SpA. The latter alternative may also explain the absence of the 48 kDa band in the 16 hour sample. The

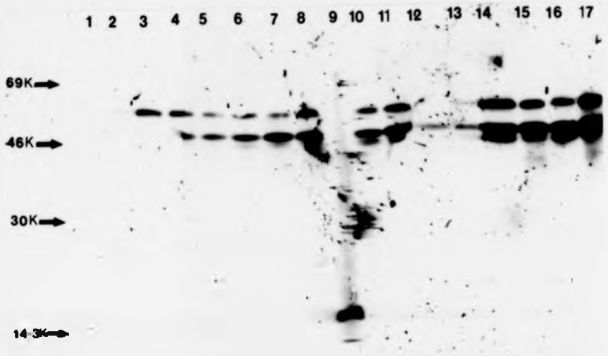


Figure 4.7 **Time Course Experiment Showing Rate of Production and Degradation of the Different IgG-binding Polypeptides Produced by JM83 [pPA31]**

Lanes contained:

(1 & 2)	1 hr post-inoculation,	(3 & 4)	2 hr post-inoculation,
(5 & 6)	3 hr post-inoculation,	(7 & 8)	4 hr post-inoculation,
(9)	molecular weight marker,	(10 & 11)	5 hr post-inoculation,
(12 & 13)	6 hr post-inoculation,	(14 & 15)	7 hr post-inoculation,
(16 & 17)	8 hr post-inoculation.		

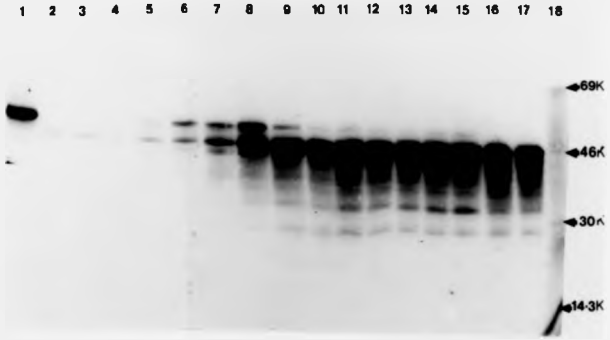


Figure 4.8 **Time Course Experiments Showing Rates of Production and Degradation for the Different IgG Binding Polypeptides Produced by JM83 [pPA34]**

Except where stated otherwise all lanes contained JM83 [pPA34] and times are post-inoculation. Lanes contained (1) *S. aureus* Cowan I, (2) 30 minutes, (3) 1 hr, (4) 2 hr, (5) 3 hr, (6) 4 hr, (7) 5 hr, (8) 6 hr, (9) 7 hr, (10) 8 hr, (11) 9 hr, (12) 10 hr, (13) 13hr, (14) 14 hr, (15) 15 hr, (16) 16 hr, (17) 16 hr cell paste and (18) molecular weight marker.

fact that definite bands are present in the 16 hour sample; (45 kDa, 43 kDa, 42 kDa, 32 kDa and 28 kDa), rather than a even spread of material along the track, and that the proportions of the various polypeptides present within the cell differ according to the time of sampling was taken, suggested that some of the cleavage activity was specific in its action. Comparison of the overall degradation patterns produced by JM83 [pPA31] and JM83 [pPA34] suggested that the full length SpA protein produced by pPA31 containing cells was slightly more resistant to degradation than the SpA protein lacking the C-terminal 69 amino acids. A similar analysis of the medium from the final sampling time for JM83 [pPA34], gave an identical banding pattern to that obtained from the cell-free lysate (Figure 4.8). Levels of SpA in the medium, prior to the decrease in SpA levels found in the cell lysates at time 15 hours, were insignificant. This suggested that the SpA present in the surrounding medium was due to cell lysis, rather than excretion.

4.2.1.9 Localisation of SpA Under Conditions of Low Expression

Protein A has been shown to be a cell wall component of *S. aureus*, covalently linked to the peptidoglycan structure, (Sjoquist *et al.*, 1972a & b), with the IgG binding regions exposed on the cell surface . As the natural host , *S. aureus*, is Gram-positive and the SpA gene has been cloned into *E. coli*, a Gram-negative organism which has a very different cell wall structure (Figure 1.2), it is not known whether the SpA would be exported to the surface of the *E. coli* cell (Chapter 1, Section 1.3). The results from the analysis of the SpA IgG-binding peptides in section 4.2.1.3 showed that the signal sequence had been correctly processed. This implied that the SpA had been transported across the cytoplasmic membrane. The apparent absence of a precursor form of the SpA molecule appeared to indicate that the processing occurs co-translationally. To clarify this situation experiments were undertaken to determine the cellular location of SpA in *E. coli*.

The localisation method used relied on a controlled disruption of the bacterial cell to yield fractions representative of a specific cellular compartment. In order to ascertain

the degree of contamination of each fraction with the other, assays for the presence of marker enzymes were undertaken. These enzymes were, alkaline phosphatase (periplasmic), glycerol-3-phosphate dehydrogenase (cytoplasmic) and NADH-oxidase (membrane). The JM83 [SpA] cells were prepared and the localisation assays undertaken as described in Section 2.2.18. If the different cellular fractions were shown to be adequately separated, then a 10 µl sample from each of the cell fractions was analysed using the standard monorocket electrophoresis conditions (Section 2.2.14.2) to quantitate the amount of SpA present, and a 200 µl sample from each of the cellular fractions examined using the Western blotting and ^{125}I labelled IgG method described in Sections 2.2.15, 2.2.16 and 2.2.17.

Localisation of SpA in JM83 [pPA30] showed SpA to be present in all the cellular fractions (Figure 4.9), however, different peptide forms were found to predominate in different cell fractions. The membrane fraction was seen to contain both the 53 kDa and 46 kDa IgG-binding peptides, with the 53 kDA peptide being the dominant form. The majority of the SpA found in the periplasmic fraction was either the 46 kDa or the 43 kDa peptide, with a small amount of the 53 kDa form being detected. All of the above three peptides were seen to be present in about the same concentration in the cytoplasmic fraction. However, the location of the marker enzymes (Table 4.5) revealed that the cytoplasmic fraction was contaminated with material from both the periplasmic fraction and the membrane fraction, to an extent that could account for the SpA levels present. As the 53 kDa SpA IgG-binding peptide is found predominantly in the membrane fraction, the smaller SpA IgG- binding peptides predominantly in the periplasm, and the N-terminus of the protein has been shown to be present on the smaller peptides, it is likely that the full size SpA peptide's association with the cell membrane is mainly due to the C-terminal section, which is deleted in the smaller IgG-binding peptides.

TABLE 4.5

Localisation of *spa*, Marker Enzymes Separation Results

JM83[Plasmid]	SpA level µg/ml	GADPH %	NADH %	Alkaline Phosphatase %
pPA30				
whole cells	2.5	-	-	-
sonicated cells	6.8	-	-	-
periplasmic fraction	5.2	1.3	0.9	87.7
cytoplasmic fraction	0.9	98.4	16.2	11.7
membrane fraction	0.5	0.3	82.9	0.6
pPA16				
whole cells	2.0	-	-	-
sonicated cells	8.5	-	-	-
periplasmic fraction	6.0	19.7	17.5	94.6
cytoplasmic fraction	1.0	80.3	10.4	2.5
membrane fraction	0.4	0.1	72.1	2.9
pPA34				
whole cells	<0.1	-	-	-
sonicated cells	<0.1	-	-	-
periplasmic fraction	<0.1	16.0	33.1	61.0
cytoplasmic fraction	<0.1	81.0	47.5	30.0
membrane fraction	<0.1	3.0	19.4	9.0

GADPH: cytoplasmic fraction marker enzyme, NADH : membrane fraction marker, Alkaline phosphatase : periplasmic fraction marker enzyme.

Localisation results for pPA31 are not shown in the above table as they were virtually identical to those obtained from pPA34.

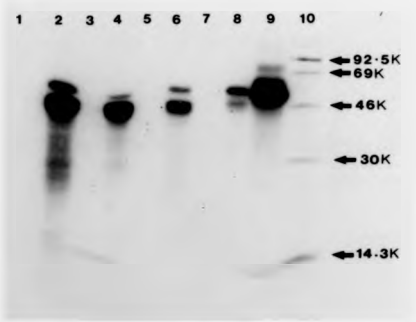


Figure 4.9 Analysis of IgG-Binding Proteins Present in the Different Cellular Compartments of JM83 [pPA16]

Lanes contained (1) JM83 [pUC8] whole cell sonicate, (2) JM83 [pPA16] whole cell sonicate (3) JM83 [pUC8] periplasmic fraction (4) JM83 [pPA16] periplasmic fraction, (5) JM83 [pUC8] cytoplasmic fraction (6) JM83 [pPA16] cytoplasmic fraction, (7) JM83 [pUC8] membrane fraction, (8) JM83 [pPA16] membrane fraction, (9) *S. aureus* Cowan I and (10) Molecular weight marker.

4.2.1.10 Localisation of SpA Under Conditions of High Expression

Localisation studies were undertaken on JM83 [pPA16], JM83 [pPA31] and JM38 [pPA34] using identical techniques to those stated for JM83 [pPA30]. These proved unsatisfactory for the following reasons. It became apparent that under low phosphate growth conditions, necessary to induce the alkaline phosphatase used as the periplasmic marker, all three normally high producing SpA plasmids synthesized little or no SpA. Thus, JM83 [pPA16] produced only 3-5% of its average yield, the SpA levels in both JM83 [pPA31] and JM83 [pPA34] were not detectable using monorocket electrophoresis. As this reduction in SpA yields was only seen in those clones producing the greatest amount of SpA, it was thought more likely to be due to loss of the resident recombinant plasmid, rather than some form of phosphate related gene regulation mechanism, as the latter would be expected to operate in both the high and low SpA production plasmids. The retention of the various SpA plasmids by JM83, when grown in either 2xYT broth or LP broth for a period of 5 hours, was confirmed by testing for ampicillin resistance (the *bla* gene of pUC8 is present in all the above SpA plasmids) by means of a comparison of the viable count on 2xYT agar plates with and without ampicillin (50 $\mu g ml^{-1}$). From the results (Table 4.6) it can be seen that loss of ampicillin resistance, indicative of plasmid loss, when grown on LP broth, explains the lack of SpA production by JM83 [pPA31] and JM83 [pPA34]. It is also of interest to note that the general level of cell viability of the latter two containing the plasmid, compared with JM83 [pPA30 and pPA16] under all growth conditions examined, was almost 10-fold lower. A similar reduction in ampicillin resistant cells was not seen to occur with JM83 [pPA16] when grown on LP broth. Therefore something other than total loss of SpA plasmid must have been responsible for the low levels of SpA produced by JM83 [pPA16] when grown in LP broth.

As a low phosphate medium affected SpA production in the high producing clones, it was thought that the normal JM83 growth medium, 2xYT broth, together with an assay for an alternative periplasmic marker such as β-lactamase, might be used. However, the levels of contamination between the different cellular compartment markers was

TABLE 4.6

Comparison of Viable Counts from JM83 [SpA] Plasmids Grown on 2xYT Broth and LP Broth With and without Ampicillin

Medium	JM83 [SpA plasmid]			
	pPA30	pPA16	pPA31	pPA34
	No. of colonies present 10^{-6}			
2xYT	480	392	86	49
2xYT (Ap 50 $\mu g\ ml^{-1}$)	457	407	60	25
LP	320	220	69	49
LP (Ap 50 $\mu g\ ml^{-1}$)	301	200	2	0

found to increase considerably in the higher SpA expression clones JM83 [pPA31 and pPA34] (Table 4.5), probably due to the increased tendency to lyse when subjected to any external stress; as reported earlier. Therefore it was decided that a localisation procedure which involved less cellular disruption would be more appropriate.

4.2.1.11 The Use of Immunological Labelling to Localise SpA Within the Cell

The recent development of immunoelectron microscopy using colloidal gold allows an alternative approach to localising SpA within the *E. coli* host. Classical cellular fractionation is not involved in the method and the need for the detection of host enzymes as markers of cross-compartmental contamination is eliminated. Therefore, the problems caused by the use of traditional studies on the JM83 [SpA] clones, such as poor cellular fractionation resulting from the fragility of cells carrying the *spa* gene when grown on a restricted medium, and the low levels of SpA produced under the physiological growth conditions necessary to induce the synthesis of alkaline phosphatase for use as the periplasmic marker) are avoided. However, immunoelectron microscopy does have its own limitations such as the need for large amounts of the protein to be present in the cell and for that protein to be in an antigenic state. The fixation and dehydration treatments necessary to prepare the cells for sectioning prior to electron microscopy (E.M) must be applied with care or they may result in an alteration of the antigenic site or in the appearance of the internal structure of the cell.

4.2.1.11.1 Analysis of *E. coli* Cell Structure Using Electron Microscopy

Preparation of cells for E.M was carried out in collaboration with A. Warnes and the Electron Microscopy undertaken by B. Dowsett.

Samples of culture for electron microscopy analysis were removed during the late exponential growth phase and prepared for E.M analysis according to Section 2.2.22. The results are shown in Figure 4.10 and 4.11. Electron lucent areas can be within seen the cells, these areas are present between the inner membrane and the outer membrane.

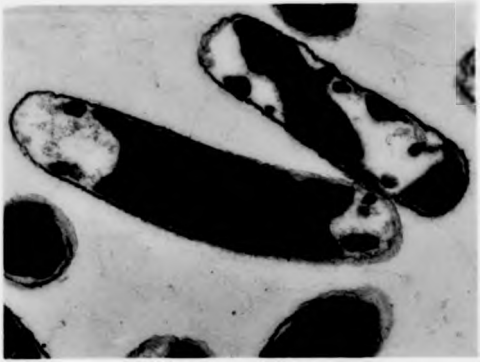


Figure 4.10 Structural analysis of *E. coli* JM83 [pPA16]

Magnification 46,750. Electron lucent areas are shown (L).

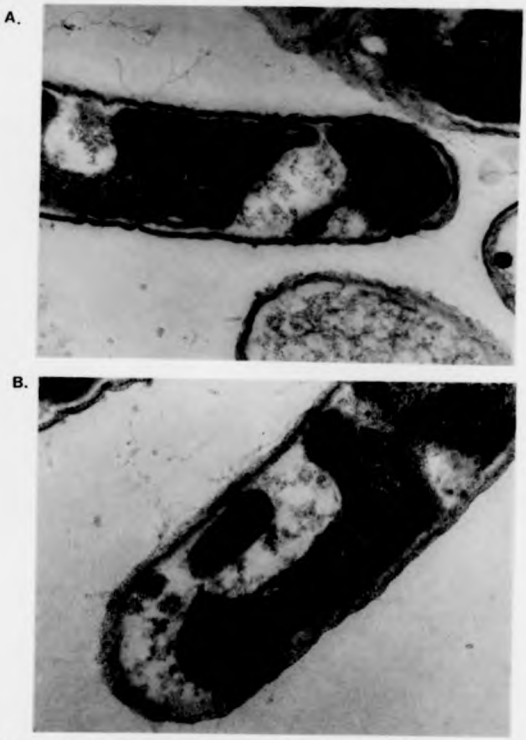


Figure 4.11 Structural analysis of *E. coli* JM83 [pPA16]

(A) magnification of 78,200, and (B) magnification of 110,400 x. Electron lucent areas are marked (L). Inner (I) and outer membranes (O) are arrowed.

4.2.1.11.2 Probing for SpA Using Immunogold Labelling

In order to establish whether immunogold labelling followed by electron microscopy would allow the location of SpA to be determined the technique was evaluated by using gold labelled antibodies to probe the surface of whole *S. aureus* Cowan I and *E. coli* JM83 [pPA16] cells. A far higher amount of gold label could be seen bound to the surface of the *S. aureus* cells than was seen in the surrounding background, ratio of approximately 10:1 (Figure 4.12(A)). The level of binding seen on the surface of *E. coli* JM83 [pUC8] cells was no higher than background levels (Warnes 1989).

The recombinant *E. coli* cells carry the plasmid pPA16 were also shown to have SpA on the surface with the gold label binding a level approximately 4-fold greater than background, but substantially lower than that seen for *S. aureus* whole cells (40% of *S. aureus* binding levels) Figure 4.12(B). Probing for SpA in JM83 [pPA16] cell sections clearly shows the immunogold localised to the electron translucent areas, Figure 4.13. The distortion of the cell walls and tears within the cell section (seen as white spots within the cells, Figure 4.13) are indicative of an increase in the section fragility caused by omitting the post-fixation osmium tetroxide step, this omission is necessary for the success of the gold immunolabelling of SpA.

4.2.2 Use of Site-Directed Mutagenesis to Specifically Alter *spa*

4.2.2.1 Rationale

A major aim of this study was to obtain enhanced expression of *spa* in *E. coli*. The measures taken to achieve this, with regards to genetic constructions, are described in detail in section 4.2.3. A number of the recombinant plasmids constructed, however, required the introduction of specific nucleotide base changes into the 5′ end of the gene. Such an undertaking is most simply achieved by site-directed mutagenesis methodology. Two specific alterations were considered appropriate:-

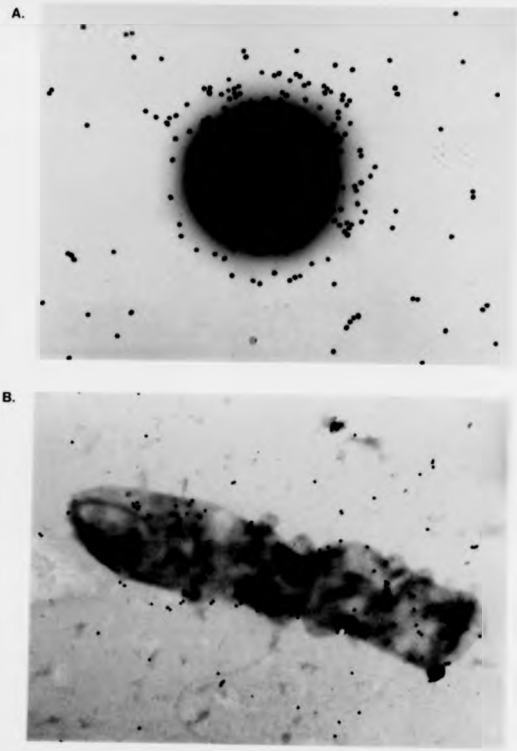


Figure 4.12 **Immunogold Labelling of the Protein A on the Surface of *S. aureus* Cowan I and *E. coli* JM83 [pPA16]**

(A) *S. aureus*, magnification of 72,000 (B) *E. coli* JM83 [pPA16], magnification of 37,400.

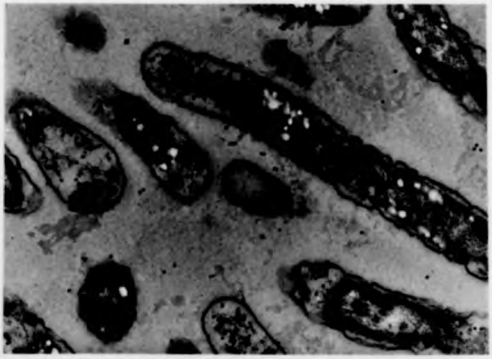
Figure 4.13 Immunogold labelling of Protein A on Sections of *E. coli* JM83 [pPA16].

The magnification levels were a; 57,200 x; b; 51,700 x and c; 35,200 x.

a.

b.

c.



i) The principle mutation change deemed worthwhile was to substitute the translational initiation codon of *spa* TTG (very uncommon in *E. coli* genes, present in only 4.5%) to ATG (the most commonly occuring initiation codon in *E. coli* genes, 91.2%). The presence of a TTG inititation codon in the *spa* gene (discussed in further detail in section 3.3.1.2) may well adversely effect its native translation intitiation frequency when expressed in *E. coli*.

ii) The experiments described in 4.2.1 did not allow any definite conclusions to be drawn as to the relative contribution of the *lac* and *spa* to the high level of expression directed by plasmid pPA34. To clarify the situation a plasmid was required which no longer carried the *spa* promoter. Removal or inactivation of the *spa* promoter would not only allow the effect of the *lacZ* promoter on *spa* production to be evaluated, it would also allow the possibility of control of SpA production during small scale fermentation (further discussed in Section 4.2.4, where continuous SpA production during fermentation can be seen to be detrimental to cell growth resulting in a decreased final yield) to be examined. The most effective way of ensuring that the *spa* promoter was non-functional was to remove the -35 and -10 transcriptional promoter sequences entirely, leaving the *spa* Shine-Dalgarno sequence intact. As there was no cleavage site for any of the commercially available restriction enzymes between the -10 promoter sequence and the Shine-Dalgarno sequence site-directed mutagenesis was employed to create one.

Oligonucleotide-directed mutagenesis is a method of *in vitro* mutagenesis which allows the construction of cloned DNA bearing a specific alteration. The oligonucleotide-directed mutagenesis technique used was based on the single primer method detailed by Carter *et al.*, 1985 (further details are given in section 2.2.12). This technique involved the chemical synthesis of an oligonucleotide consisting of a mutant sequence, which is hybridised to its complementary wild-type DNA (prepared as a single-stranded template using an M13 derivived vector) forming a mutant-wildtype duplex. The

oligonucleotide mutant once hybrdised serves as a primer for *in vitro* enzymatic DNA synthesis of the regions which are to remain genotypically wild-type. A double-stranded duplex is formed, which is subsequently segregated *in vivo* into separate mutant and wild-type clones, which can be detected by screening with the radio-labelled mutagenic oligonucleotide.

4.2.2.2 Oligonucleotide Design

The longer the oligonucleotide is, the more likely that it will recognise only the target sequence within the *spa* gene. Although it is now easily possible to produce synthetic oligonucleotides of 50-100 bases in length, these are not particularly suitable for site-directed mutagenesis for two reasons : firstly the chances of an error in the synthesised sequence increases with length and secondly the screening method used for detecting the mutants relies on differential annealing procedure involving radiolabelled mutagenic oligonucleotide. This procedure is dependent on the difference in the dissociation temperature (Td) of a perfectly matched oligonucleotide duplex and that of a mismatched duplex. It is possible to calculate the Td of a duplex by using the values stated in the Wallace rule : G - C base pairing = 4 C and A - T base pairing = 2 C (Wallace *et al*, 1979). However the relationship between oligonucleotide length and Td is nonlinear, and it is generally accepted (Smith 1986) that this rule become progressively more invalid as oligonucleotides become greater in length than 20 bp. Therefore it was decided to design mutagenic oligonucleotides consisting of 20 nucleotides.

The strictest criterion regarding the placement of the mismatch within the oligonucleotide is that it must not be placed within 4 nucleotides of the 3′ end, as this might result in repair of the mismatch by the 3′ > 5′ exonuclease activity present in the DNA polymerase Klenow fragment. The presence of 3 nucleotides following the mismatch has been shown to protect it from this exonuclease activity (Gillam and Smith, 1979). Placement of the mismatch close to the centre of the oligonucleotide, is also important if the oligonucleotide is to be used as a probe to screen for the mutant.

Szostak *et al* (1980) have demonstrated that the central positioning of the mismatch yielded the greatest binding differential between a perfectly matched duplex and a mismatched duplex.

The final step in designing an oligonucleotide is to analyse the DNA sequence of both the vector, in this case M13mp8, and the rest of the cloned fragment for regions of partial complementarity and so identify any competing sites from which priming might also occur. There should be as little dyad symmetry as posssible in the oligonucleotide in order to advoid the possibility of self-hybridisation occurring.

Observing the above criteria and using the comparison programs supplied by DNASTAR Inc (Madison, USA) set to show regions of 50% or greater homology within a 20 nucleotide window; a range of possible mutagenic oligonucleotides 20bp in length (with the mismatch occupying a central position, at least 5bp from either the 5′ or 3′ end) were analysed.

The DNA sequence for the oligonucleotide designed to introduce an *Eco*R1 site between the *spa* -10 promoter sequence and the S.D sequence by altering the sequence AAATAC to GAATTC, whilst consisting of a static 13 nucleotide core region could be extended by up to 7 nucleotide in either the 5′ or 3′ direction, as shown in Figure 4.14. Computer analysis showed no homologies of 50% or greater existed between any of the potential mutagenic oligonucleotide sequences and the vector M13mp8. However all the potential mutagenic oligonucleotides showed many homologies of 60% and higher when compared to the rest of the cloned spa fragment. Taking into account the fact that a predominance of matches at the 3′ end of the oligonucleotide will make a more efficient priming site than the same degree of complementation at the 5′ end (Zoller *et al*, 1983) the nucleotide sequence indicated in Figure 4.14 was selected to be used as a mutagenic oligonucleotide. For, although it had one match of 68% homology with a region further downstream in the *spa* gene only one of the 6

Figure 4.14 ***spa*** **DNA Sequences to which the complementary sequences for the two oligonucleotides for site directed mutagenesis were synthesized**

A) Oligonucleotide to introduce an *Eco*RI between the -10 promoter and the Shine-Delgarno sequence.

A A
gatgactTTACGAATTCATAcaggggg

B) Oligonucleotide to alter the translation initiation codon from TTG to ATG.

T
acaggggtatTAATATGAAaaagaaaaaca

The oligonucleotide is complementary to the DNA sequence shown. The difference between the oligonucleotide sequences and the wild-type *spa* DNA sequence which will creat point mutations, are indicated: superscript letters represent the nucleotides in the original *spa* nucleotide sequence which were altered. Upper case letters indicate which nucleotides should be included in the synthetic oligonucleotide. Lower case letters indicate the range (in both the 3' and 5' direction) from which a 20 bp oligonucleotide could be selected (following the rules given on Section 4.2.2.2). Oligonucleotides synthesized for mutagenesis are indicated by underlining.

bases at the 3′ end of the oligonucleotide was involved, thus the oligonucleotide was unlikely to hybridize to this competing region.

A similar analysis of the vector and *spa* DNA sequence for regions complementary to the possible designs of oligonucleotides to alter the *spa* translation initiation codon from TTG to ATG gave many matches of 65% homology and over against both M13mp8 and the non-target region of the *spa* insert. The oligonucleotide with the lowest degree of homology to the latter two regions was selected for the mutagenic oligonucleotide, see Figure 4.14. This oligonucleotide gave 16 homologies of 60% and above with M13mp8, including 7 matches with the 5 bases at the 3′ end having 100% homology. The oligonucleotide also showed 2 matches of over 65% homology with non-target regions of the *spa* gene, one of these exhibiting 100% homology for the 6 bases at the 3′ end of the oligonucleotide.

Several regions of the *spa* gene were shown to exhibit a high degree of homology to one or other of the two nucleotide probes. The specificity of the oligonucleotide for the correct target sequence was checked by using the oligonucleotide sequence as a dideoxy-sequencing primer for single-stranded M13 mp8 DNA containing the spa gene (Figure 4.15). Although both synthetic oligonucleotides bound to the target spa DNA sequence and initiated DNA synthesis, the resultant sequence "ladders" became unreadable approximately 21-26 nucleotides ustream of the *Dra*1 site. Comparison of the oligonucleotide sequence with the *spa* sequence at these points (Figure 4.16) showed that the appearance of multiple bands on the autoradiograph coincided with areas of strong homology. These extra bands are probably caused by multiple hybridisation resulting from the use of these particular synthetic oligonucleotides as primers, as no interference was seen on the original sequencing of the spa gene using M13 universal primer. The overall level of homology between the oligonucleotides and thse non-specific regions was 50% over a distance of 20bp (although at the point the extra banding occurs; X and Y (Figure 4.15 and 4.16) the level of homology was 83% over a distance of 12 nucleotides and 89% over a distance of 9 nucleotides); and as such

Figure 4.15 Dideoxy-sequencing Showing Multiple Binding of the Mutagenic Oligonucleotides

Figure 4.16 **Examination of Possible Hybridisation Between the Synthetic Oligonucleotide Sequences with Non-target DNA 5' to the *spa* Gene**

Homology X

```
ACAAAATTTTTATTTTATAA
    ||||  |||||  |
GACTTTACAAATACATACAG
```

50% homology over 20 bp

Homology Y

```
ATCTTTGGACAAAATTTTTA
  |   |     |||  |||||
GGGGTATTAATTTGAAAAAG
```

45% homology over 20 bp

The upper sequence of each homology pair is native *spa* DNA, while the lower sequence is the synthetic oligonucleotide (with X = *Eco*R1 mutation and Y = ATG initiation codon mutation). The position of *spa* DNA sequence with respect to the *Dra*I site and target priming site are shown on Figure 4.15.

would not have been detected as significant within the limits of the homology search undertaken using the computer (homologies of 60% and greater over a window of 20bp), therefore, a similar or more extensive effect may occur in the regions of greater homology shown to exist from the computer search.

4.2.2.3 Site-Directed Mutagenesis

During site-directed mutagenesis it is commonplace for additional mutations to those intended to be introduced into the target DNA. It is, therefore essential that the authenticity of mutated DNA be thoroughly checked prior to subsequent subcloning into a plasmid vector. To avoid having to resequence the entire *spa* gene after oligonucleotide mutagenesis only the part of the SpA gene containing the desired mutagenic changes was isolated, and an equivalent region in the wild-type, plasmid borne, *spa* was then replaced with this mutagenic subfragment. The *spa* gene was subjected to SDM and the mutated region subcloned as a 223 bp *DraI*/*MaeI* restriction fragment (see Figure 4.17).

Mutagenesis was carried out as detailed in section 2.2.12. Detection of the *spa* sequence containing the mutations was dependent on differentiation between the Td values for the hybridisation between the γ^{32}P labelled mutagenic probe and the single stranded template DNA, the latter being either the wild-type spa or the spa gene containing the mutagenic alterations. The Td values of the different mutations and wild-type regions are indicated below:

	Td value ^{o}C	
	wild-type *spa* - mutagenic probe	mutant *spa* - mutagenic probe
*Eco*R1 mutation	43	49
ATG start codon	45	47

Identification of the mutant spa DNA was undertaken as follows. Excess ^{32}P oligonucleotide probe was removed from the filters by 3 washes in 6 x SSC at 37 C and the filters autoradiographed for an hour, Figure 4.18. The filters were then washed at 45 C for 1 minute. The radioactivity from the filter containing *Eco*R1 mutation was seen to drop from 200 counts/min to 50 counts/min whereas no such fall in radioactivity was observed for the filter containing the ATG mutation, both filters were autoradiographed for 3 hours. The presence of the *Eco*R1 mutation was easily identifiable, Figure 4.18(A), no differentiation between probe intensity was visible for the ATG mutation filter, therefore it was subjected to a set of 1 minute duration 6 x SSC washes at gradually increasing temperatures until a substantial drop in radioactivity levels was observed at 48^{O}C. The filter was then autoradiographed for 3 hours, Figure 4.18(A). Four "positive" colonies were then plaque purified and four templates prepared from each, Section 2.2.12.4. The mutagenic alteration was checked by dideoxy-sequencing using an oligonucleotide primer complementary to a region in the signal sequence 45 bp downstream of the *Mae*1 site, Figure 4.17. The entire DNA sequence of the *Mae*1 - *Dra*1 fragment, was resequenced using this primer to ensure that no other mutagenic alterations had occured. Although the plaques from the *Eco*R1 mutation gave a higher percentage of mutants (60%) than the ATG mutation (40%), the use of the *Eco*R1 oligonucleotide also resulted in more incorrect mutations than the use of the ATG oligonucleotide (with 15% of the *Eco*R1 mutant templates containing multiple copies of the oligonucleotide). However because the *Eco*R1 mutation had been far easier to identify than the ATG mutation, the *spa* clone containing the both the *Eco*R1 mutation and the ATG mutation was constructed by carrying out a further mutagenesis step with the *Eco*R1 oligonucleotide on one of the successful ATG mutated templates. Detection of a successful mutant was carried out as previously decribed for the *Eco*R1 mutation, Figure 4.18.

4.2.2.4 Incorporation of Mutagenic Changes into Recombinant Plasmids

In order to be able to carry out a direct comparison of the effect of the mutagenic changes on the levels of SpA production in plasmids carrying either the full length *spa*

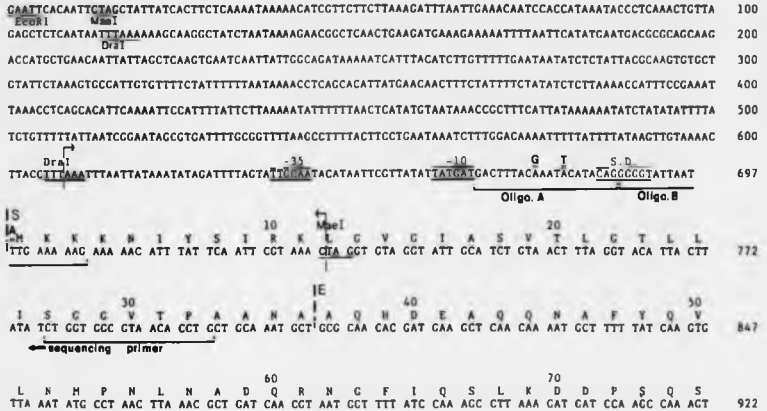


Figure 4.17 ***DraI-MaeI*** **Fragment Used for Sub-cloning into Wild-type *spa* Following Oligonucleotide Mutagenesis**

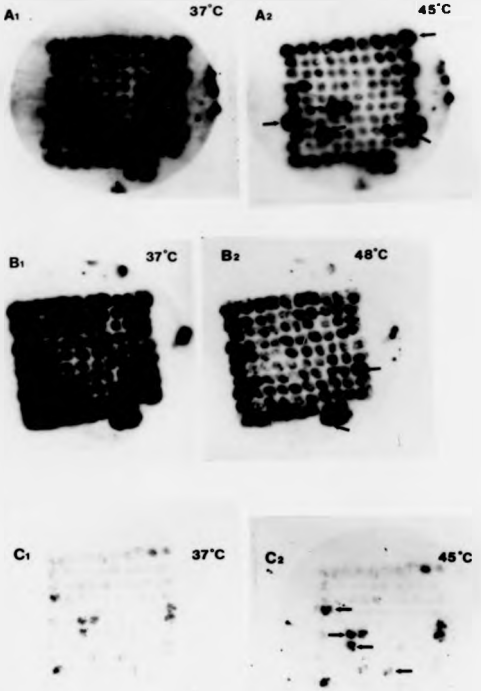


Figure 4.18 **Detection of Positive Mutants Following the *Eco*RI and ATG Oligonucleotide Mutagenesis**

A: *Eco*RI site mutagenesis
B: ATG start codon mutagenesis
C: *Eco*RI site mutagenesis carried out on earlier positive ATG mutagenic clone

gene (pPA30 and pPA31) or the truncated C-terminal version of the *spa* gene (pPA16 and pPA34) it was decided to produce both the above versions with an identical series of mutagenic changes (Table 4.7).

Recombination of the mutated regions of the spa DNA with the "complementary" non mutated *spa* DNA to produce the complete *spa* gene required a series of DNA manipulations. Non-mutated *spa* DNA was isolated from a large scale plasmid preparation of pPA31. The mutated *spa* DNAs were isolated from large scale preparations of the replicative form of phage DNA, as described in Section 2.2.4.4. The series of manipulations necessary to construct *spa* plasmid pPA37, pPA38, pPA39, pPA40, pPA41 and pPA42 are detailed in Figure 4.19.

The transformation mixes were plated on to minimal agar containing Ap50 and BCIG (and IPTG for the JM101 transformants), to allow the detection of those carrying *spa* inserts. Twenty positive (white) colonies from each of the different plasmid types were selected and the DNA extracted for restriction enzyme mapping by the method of Birnboim and Doly (1979). Of the colonies examined for pPA37, pPA38, pPA41 and pPA42 63% contained the *spa* gene insert, and all but one gave the predicted restriction enzyme pattern. Only 35% of the pPA39 colonies and 10% of the pPA40 colonies contained an *spa* gene fragment, half of which had no *Dra*I or *Mae*I cleavage site.

Plasmid sequencing (as described in Section 4.2.2.11) was undertaken on 4 of the *spa* containing plasmids from each group. The DNA sequencing gels produced were of very poor quality, and despite alterations in both the concentrations of the sequencing chemicals and the use of different plasmid preparations, no improvement in the subsequent DNA sequencing was seen. Even though the nucleotide composition of all of the recombined regions of the plasmids could not be determined, it could be seen whether the *spa* fragments and pUC8 vector had recombined in the correct order. The pPA39 and pPA40 plasmids which had shown the aberrant restriction pattern were

TABLE 4.7

Summary of the Difference Between the Six Mutagenic *spa* Plasmids

C-terminus comparable to earlier plasmid no:	Mutagenic alteration		
	***Eco*RI site**	**ATG initiation codon**	***Eco* RI site + ATG initiation codon**
pPA30/31	pPA37	pPA39	pPA41
pPA16/34	pPA38	pPA40	pPA42

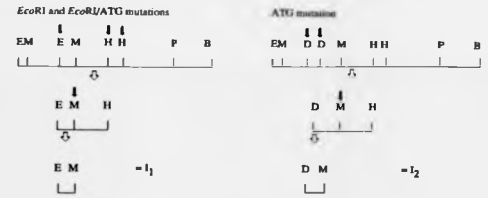
The mutagenic *Eco*RI cleavage recognition sequence was introduced between the predicted -10 promoter (Pribnow box) and the Shine-Dalgarno sequence, to produce a possible cleavage site between nucleotide positions 675 and 676.

The ATG translation initiation codon replaces the natural TTG translation initiation codon of the *spa* gene at amino acid position 1.

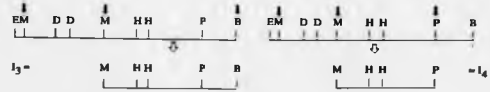
Nucleotide and amino acid numbering is according to Figure 3.2

Figure 4.19 **Construction of plasmids pPA37, pPA38, pPA39, pPA40, pPA41 and pPA42**

A) SpA gene:



B) Original *spa* gene



C) Recombination of the SpA fragments

$I_1 + I_3$	=	pPA37
$I_1 + I_4$	=	pPA38
$I_2 + I_3$	=	pPA39
$I_2 + I_4$	=	pPA40
I_1 (ATG) + I_3	=	pPA41
I_1 (ATG) + I_4	=	pPA42

Legend: Both M13mp7 and pUC8 contain multiple sites for the restriction enzymes *Mae*1 and *Dra*1. Fragmentation of the vector and *spa* insert with these enzymes would result in several vector DNA fragments of a similar size (indistinguishable using agarose gel isolation) to the required *spa* fragment. Therefore the *spa* gene was initially isolated as an *Eco*R1-*Bam*H1 fragment (A and B)

The following annotation for restriction endonuclease sites has been used; E: *Eco*R1, M: *Mae*1, H: *Hin*dIII, P: *Pst*1, D: *Dra*1 and B: *Bam*H1. Cleavage of a restriction site is indicated by the presence of an arrow (⇩) above the sites (A and B).

The *spa* fragments (represented as Inserts, 1, I>4) were then recombined with one another as shown in C (above) and ligated into appropriately cleaved pUC8 vector, which was then transformed into *E. coli* strains JM83 and JM101.

found to lack the *Dra*1-*Mae*1 *spa* fragment entirely.

4.2.3 Examination of Factors Effecting SpA Expression

The initial strategy adopted to obtain high expression of protein A in *E. coli* was to position the gene on a high copy number plasmid (pUC8) 3′ to the promoter of the *E. coli lac* operon. The *lac* operon of *E. coli* consists of six different regions, which can be subdivided into two sections. One, which controls the level of expression of the rest of the operon, consists of three regulatory features; the sequence coding for the repressor gene I, the promoter region containing both the cAMP-CRP site and the Pribnow box, and the operator region containing the repressor protein binding region and the ribosome binding site. The second codes for the three structural genes: β-galactosidase (Z) which hydrolyses lactose to glucose plus galactose and is therefore essential for lactose catabolism, galactosidase permease (Y) which transports lactose and many related substances into the cell, and β-galactosidase transacetylase (A) whose exact physiological roles has not yet been identified. The *lac* operon is subject to both repression control and activation control (Figure 4.20).

The plasmid pUC8 contains only a small region of the *lac* operon: the promoter region (P); the operator region (O), containing the repressor protein (I) binding site; and the first 59 codons of the *lacZ* structural gene, with the multiple cloning site (consisting of 11 amino acids) inserted between the sixth and seventh codon of the *lacZ* structural gene. Although pUC8 does not code for the *lac* repressor protein (I), it does contain the repressor binding region. Therefore, the function of the *lac* promoter of pUC8 can be controlled by altering the amount of repressor present in the host system in proportion to the number of copies of the *lac* gene present.

In the case of the initial recombinant plasmids constructed, expression data described in section 4.2.1 appeared to indicate that the presence of the *lac* promoter 5′ to the *spa* structural gene (pPA16) had no beneficial effect on the level of SpA produced by *E. coli*, unless a region of DNA located between the *lac* promoter and the *spa* promoter

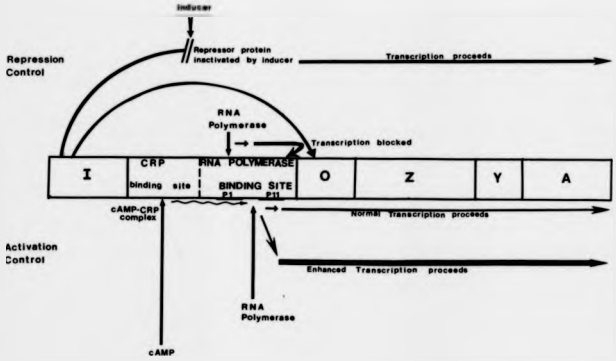


Figure 4.28 Control Mechanisms of the *lac* Operon

The *lac* operon has two control mechanisms that operate at the level of transcription.

Repression control: The repressor protein (I) is an allosteric protein existing in two forms, only one of which binds to the operator region. In the absence of an inducer molecule it will favour the operator-binding conformation and blocks access of the RNA polymerase to the promoter, thus preventing transcription of the *lac* structural genes. In the presence of inducer, which will bind to the repressor and cause it to assume a operator-non-binding conformation, the RNA polymerase has access to the promoter and initiates transcription of the structural genes

Activation control: This is dependent on the levels of cAMP which show an inverse correlation with the cell growth rate. The CRP gene (unlinked to the *lac* operon) codes for the cyclic 3 , 5 -AMP receptor protein. The CRP protein is also allosteric in nature, in one conformation (caused by the presence of cAMP) it recognises the CRP DNA binding site, in the absence of cAMP it does not. The cAMP-CRP complex will bind to the CRP-binding site adjacent to that for RNA polymerase and in doing so enhances the affinity of the promoter for RNA polymerase and so increases transcription.

was first deleted (pPA34). It was, therefore of interest to determine whether the high level of *spa* expression directed by plasmid pPA34 was due either to the *lac* promoter and/or the *spa* promoter.

4.2.3.1 Repression of *lac* using *lacI*^{q} at low gene dosage

The earlier expression work was carried out using *E. coli* JM83 as the host for the recombinant based on pUC8. This strain has the *lac-pro*AB region of the chromosome deleted, and replaced by the *lacZ* Δ M15 mutation (which will produce a structural *lacZ* protein that complements that produced by pUC8). The absence of the *lacI* gene in JM83 results in the constitutive expression of the *lac* promoter. The fact that transcription from *lac* cannot be repressed in JM83 meant that the extent to which *lac* and the *spa* promoter itself contributed to the expression of the *spa* gene, could not be ascertained. Use was therefore made of an *E. coli* strain which contained a *lacI* gene as a host for the SpA plasmids. JM101 was selected as this was very similar to JM83, and carried the same chromosomal (*lac-pro*AB) deletion, however it differed from JM83 in that it carried a copy of *lacI*^{q} and *lacZ* M15 on an F' episome. The superscript 'q' of *lacI*^{q} indicates that this particular gene contains an "up" promoter mutation which causes a ten-fold increase in the expression of the repressor protein (Muller-Hill *et al.*, 1968, 1971). As the repressor is present the *lac* promoter of any pUC8-based any plasmid transformed in JM101 will be inducible rather than constitutive.

Cells of JM101 carrying either pPA16, pPA31 or pPA34 were cultured in 10 ml 2xYT (ampicillin 50 μg ml^{-1}) at 37^{o}C overnight, in the absence or presence of the *lac* inducer IPTG (final concentration 1 mM). The SpA was recovered from the cells and quantitated as detailed in Section 2.2.13.2 and 2.2.14.2. From the results presented in Table 4.8 for pPA31 and pPA34 the level of SpA increased substantially in the presence of the inducer IPTG, indicating that the *lac* promoter of pUC8 is apparently contributing to *spa* expression. The lack of a similar increase for pPA16 would appear

TABLE 4.8

De-repression Of the *lac* Promoter Of The SpA Plasmids In JM101 Using IPTG

Plasmid	SpA levels ugmg^{-1} cells dry wt.		
	-IPTG	+IPTG	% increase in SpA in presence of IPTG
pPA16	17.7 (1.9)	18.1 (1.2)	2.1
pPA31	43.7 (9.7)	51.4 (8.7)	17.6
pPA34	72.5 (16.1)	86.4 (21.6)	19.2

Samples from 6 sets of experiments.

std. deviation values in brackets (n-1)

to be consistent with the hypothesis that the 609 *S. aureus* nucleotides prior to the *spa* promoter interfere with transcriptional readthrough from the *lac* promoter. It was also apparent that high levels of SpA were produced in the absence of inducer. There were two possible explanations for this. Firstly, the *spa* promoter itself could be responsible for the majority of *spa* expression. Or secondly, this "background" production of SpA can be accounted for by the fact that the equivalent number of operator sequences are present within the cell to the copy number of the plasmid. In the case of pUC8 the copy number has been shown to be between 500 and 600 copies per cell. In contrast, the *lacI*^{q} gene is only present at 2 to 3 copies per cell i.e., the copy number of the F′ factor. Thus the excess of operator sequences could effectively titrate out all the available repressor protein molecules, with the result that expression from the *lac* promoter elements of a high percentage of the plasmid population would be constitutive.

4.2.3.2 Repression of *lac* using *lacI*^{q} at high gene dosage

One way to overcome the inbalance between *lac* operator sites and repressor protein would be to increase the levels of the latter by locating the *lacI*^{q} gene to a plasmid with a higher copy number than F′ . The plasmid pNM52 was especially constructed for this purpose (Gilbert *et al.*, 1986). It was constructed by inserting the *lacI*^{q} gene into pACYC184, and is therefore compatible with any plasmid derived from pUC8. It has a copy number of 20 copies per cell, and therefore is theoretically able to direct the expression of up to 10- fold higher levels of repressor protein than the F' episome of JM101.

Plasmid pNM52 was transformed into JM83 [pPA16], JM83 [pPA31] and JM83 [pPA34] and the cells were cultured either with or without 1mM IPTG present. Contrary to what had been expected, not only did the addition of pNM52 fail to repress the production of SpA, it actually appeared to increase the levels, (Table 4.9). The addition of the inducer IPTG further increased the level of SpA recovered from the cells. To ascertain that the above anomaly was not restricted to the JM83 [SpA] system, pNM52 was then transformed into MC1061 [pPA31] and MC1061 [pPA34],

TABLE 4.9

Controls Of The *lac* Promoter in the *spa* Plasmids by Use of *lacI*^{q} Expression From pNM52

Host (plasmid)	SpA Levels μg ml^{-1} Cells Dry Wt.		
	-pNM52	+pNM52	+pNM52 +IPTG
JM83 [pPA16]	32.2	36.0	42.61
	(7.8)	(7.6)	(9.2)
JM83 [pPA31]	65.6	89.0	106.3
	(13.2)	(20.0)	(16.9)
MC1061 [pPA31]	57.3	80.2	90.7
	(6.4)	(14.9)	(5.9)
JM83 [pPA34]	123.7	170.8	201.2
	(15.1)	(29.0)	(28.0)
MC1061 [pPA34]	86.2	95.6	112.9
	(16.1)	(29.0)	35.6)

SpA levels were not affected by adding IPTG to control cultures containing the SpA plasmid but not pNM52. std. deviation values are in brackets.

where exactly the same effect was seen to occur (Table 4.9). It was of interest to note that cells containing both an SpA plasmid and pNM52 were extremely fragile, cultures to be grown could only be inoculated from a -70^{o}C stock as colonies on agar plates were found to lyse within 24 hours, when grown at 37^{o}C.

4.2.3.3 Deletion of the *spa* promoter

Data obtained from the experiments utilising *lacI*^{q} to regulate *spa* expression, described in two previous sections (4.2.3.1 and 4.2.3.2), did not allow any conclusions to be drawn as to the reason for the high level of SpA production in the absence of IPTG. As one possible source of transcription could be the *spa* promoter itself, recombinant plasmids were constructed which no longer carried the *spa* promoter. The construction of these plasmids (pPA37 and pPA38) enabled a direct comparison with the SpA levels produced by pPA31 and pPA34, as they are identical with regards to the *spa* encoding DNA, but lack the *spa* promoter region.

Cells of *E. coli* JM83 carrying pPA31, pPA34, pPA37 or pPA38 were cultured in 500 ml quantities, harvested and assayed for SpA production as detailed in section 4.2.1.2. The graph of growth against SpA production (Figure 4.21) for JM83 carrying the plasmid pPA38, closely resembled those of pPA34. In contrast, a similar comparison between pPA37 and pPA31 demonstrated that SpA production per mg of cells was higher during the early period of growth in cells possessing the former plasmid, though the final levels of SpA per mg of cells from an overnight culture of both clones were the same. Similar experiments were undertaken using JM101 in place of JM83 as the host strain. Cultures were grown in both the presence and absence of 1mM IPTG and harvested at 9 hours post-inoculation. SpA yields of both the JM83 and JM101 clones were determined and the results analysed as described for pPA16, pPA31 and pPA34 in section 4.2.1.2., the results are presented in Table 4.10.

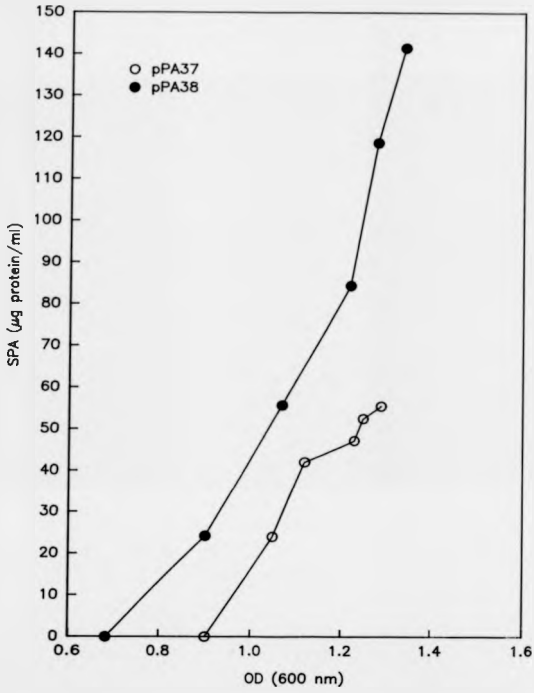


Figure 4.21 The effect on SPA production of removal of the spa promoter from JM83[pPA31] & [pPA34] to give plasmids [pPA37] & [pPA38] respectively

TABLE 4.10

The Effect on SpA Production of the *lac* Promoter Under Repressed and Derepressed Growth Conditions in the Absence of the *spa* Promoter

Host	Plasmid present	
	pPA37	pPA38
JM83	118.5	244.5
	(4.3)	(31.6)
JM101 no IPTG	85.2	253.3
	(9.1)	(21.1)
JM101 + 1 mm IPTG	101.6	328.8
	(6.7)	(46.0)
JM101 % increase in SpA yields in the presence of IPTG	19.2	29.8

std. deviation values are shown in brackets.

JM83 results are taken from 5 sets of experiments.

JM101 results are taken from 3 sets of experiments.

Overall, the results with the host JM83 clearly demonstrate that removal of the *spa* promoter has no effect on the levels of SpA obtained. Thus the level of *spa* expression directed by pPA37 and pPA38 are the same as that directed by pPA31 and pPA34, respectively. The only significant difference in the SpA levels seen was between the plasmid possessing the entire *spa* gene (pPA37) and pPA38, which carries the *spa* gene truncated at the C-terminus. The latter plasmid directed the expression of twice the levels of SpA as the former. The same pattern was previously observed with the corresponding progenitor plasmids pPA31 and pPA34 (Table 4.1).

Similar expression data was obtained when JM101 was employed as the host organism, with the exception that cells carrying pPA38 produced significantly higher levels of SpA when grown in the presence of 1mM IPTG. The high standard deviation value for this result does suggest that this may be a spurious observation. A direct comparison between the SpA levels of the mutagenic clones in JM101 with the SpA levels of pPA31 and pPA34 in JM101 was not possible, as the former were grown in 500 ml cultures and the latter in 10 ml cultures. The use of these two different culture volumes with JM83 clones has previously resulted in a two-fold difference in SpA production (Table 4.1). The percentage increase in SpA production by the mutagenic plasmids (Table 4.10) upon derepression of the *lac* promoter by the presence of IPTG in the growth media is consistent with that observed for pPA31 and pPA34 (Table 4.7), except for pPA38 which shows the higher increase in SpA production of 29.8%.

4.2.3.4 Removal of *spa* from the transcriptional control of the *lac* promoter.

The above results suggest that the *lac* promoter was solely responsible for the high expression of SpA in vectors such as pPA34. The fact that expression of *spa* was obtained in *E. coli* using pPA12 (4.2.1) demonstrates that the *spa* promoter can function in *E. coli*. To date, however, the efficiency of the *spa* promoter on a high copy number plasmid could not be estimated due to the presence of the "upstream" *lac* promoter. The simplest way to ensure that the *lac* promoter was not contributing to the

transcription of the *spa* gene, whilst retaining it in an otherwise identical multicopy plasmid, was to remove the *lac* promoter from the pPA plasmids. The region encompassing the *lac* promoter was removed from pPA16, pPA31 and pPA34 without altering the *spa* β-galactosidase alignment to form plasmids pPA44, pPA45 and pPA46, equivalent to the pUC8 constructs pPA16, pPA31 and pPA34; respectively.

To remove the *lac* promoter, the *spa* plasmids were cleaved with *Eco*RI and *Sca*I, and the *Eco*RI termini blunt-ended using T_4 polymerase. The blunt-ended *Eco*RI-*Sca*I fragment containing the *spa* gene was then isolated using agarose gel electrophoresis and the DNA recovered as detailed in sections 2.2.4.6 and 2.2.4.7, pUC8 DNA was cleaved with *Sca*I and *Pvu*II and the 5′ termini of the fragments dephosphorylated to prevent any later self ligation of the blunt ends. The three pUC8 DNA fragments were separated on a 1% agarose gel and the 1.55 kb *Sca*I-*Pvu*II fragment recovered and ligated with the blunt ended *Eco*RI-*Sca*I [*spa*/pUC8] fragment (Figure 4.22). *E. coli* JM83 was transformed with the ligation mix and plated onto a selective media containing 50μg ml^{-1} ampicillin. The two fragments could ligate in either orientation but only the required ligation would result in the correct reconstruction of the *bla* gene. Ten colonies from each transformation were streaked out on 2xYT, ampicillin (50 μgml^{-1}) plates incubated at 38°C overnight and the DNA isolated using the STET procedure (Methods Section 2.2.4.2.1). The size of the *spa* insert was then checked by cleavage with *Rsa*I.

JM83 cells containing either the above *lac* promoter-less plasmids or pPA16, pPA31 and pPA34 (as controls) were grown overnight at 37°C in 10 ml of 2xYT ampicillin (50 μg ml^{-1}). A 2.5 ml aliquot of these overnight cultures were used to inoculate 500 ml of prewarmed 2xYT (50 μg ml^{-1}) broth in a 2 litre flask, which were shaken at 200 rpm at 37°C. Ten ml samples were taken at hourly intervals from 2 hours post inoculation to 9 hours post inoculation, a final sample of both cells and supernatant was taken at 24 hours post inoculation. The SpA content of the samples was determined as previously described in Section 2.2.14.2.

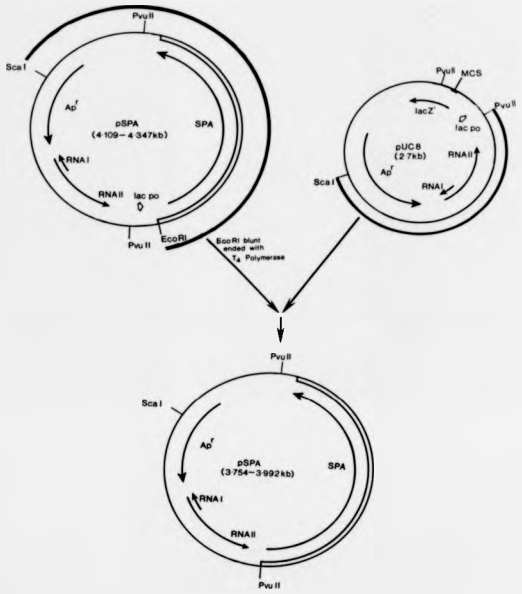


Figure 4.22 **Construction of the *lac*⁻ Plasmids pPA44, pPA45 and pPA46**

No SpA was detected in any of the 24 hour post inoculation sample supernatants. Growth curves showing μg ml^{-1} SpA against culture density were plotted for all the above cultures (Figure 4.23). The removal of the *lac* promoter's contribution towards transcription resulted in a substantial reduction in the amount of SpA recovered from the cultures. Further 24 hour samples were taken in a series of repeat experiments and the average SpA yield determined for each plasmid type (Table 4.11). Those plasmids where SpA expression is entirely dependent on the *spa* promoter alone produced less than half (44 - 49%) of the amount of SpA obtained from the equivalent plasmid which carries both the *spa* promoter and the *lac* promoter in tandem.

4.2.3.5 Alteration of the *spa* translational initiation codon

The site-directed mutagenesis experiments described in 4.2.2 had resulted in the replacement of the *spa* TTG start codon with ATG. This mutation change was effectively introduced into the plasmids pPA30, pPA34, pPA37 and pPA38 to yield the plasmids pPA39, pPA40, pPA41 and pPA42, respectively. Cells carrying all 8 of these plasmids were separately cultivated and the level of SpA produced estimated as previously described.

Alteration of the translation initiation codon from TTG to ATG has no effect on the yields of SpA obtained from JM83 and JM101 or on their rate of growth (Figure 4.24). Removal of the DNA upstream of the *spa* SD sequence including the *spa* -10 and -35 promoter region also appears to have little effect on the levels of SpA expressed, with the derepression of *lac* promoter in JM101 resulting in the same order of increase in SpA production regardless of the presence of the *spa* promoter (Table 4.12). Thus, the presence of the *lac* promoter in the SpA plasmids can account for all of the SpA produced and the additional presence of the *spa* promoter in some of the plasmids would seem to either not interfere with transcriptional readthrough from the *lac* promoter or alternatively compensates for any interference by its own activity.

TABLE 4.11

Comparison of Maximum SPA Yields of the *lac* Promoter-Plasmids with the Comparable *lac* Promoter + Plasmids

µg SPA / mg cells dry wt		% difference
pUC8 [lac +]	pUC8 [lac -]	A
pPA16:55.7	pPA44:25.8	46.3
pPA31:124.0	pPA45:50.1	40.4
pPA34:293.9	pPA46:132.4	45.0

TABLE 4.12

Effect on SpA Production of i. Alteration of the *spa* Translational Initiation codon [pPA39] and [pPA40] ii as i plus the *lac* Promoter Under Repressed and Depressed Growth conditions in the Absence of the *spa* Promoter [pPA41] and [pPA42]

Host	Plasmid Present			
	pPA39	pPA40	pPA41	pPA42
	SpA $\mu g mg^{-1}$ Cells Dry Weight			
JM83	122.3	249.9	121.0	258.1
	(10.4)	(33.2)	(11.3)	(40.1)
JM101	97.1	241.2	121.1	204.5
no IPTG	(15.1)	(10.2)	(16.0)	(30.1)
JM101	177.7	278.1	143.7	237.1
+1mM IPTG	(18.4)	(29.2)	(20.1)	(36.3)
JM101	21.2	15.3	18.7	15.9
% increase in SpA yields in the presence of IPTG				

std. evaluation values are shown in brackets.

JM83 results are taken from 5 sets of experiments.

JM101 results are taken from 3 sets of experiments.

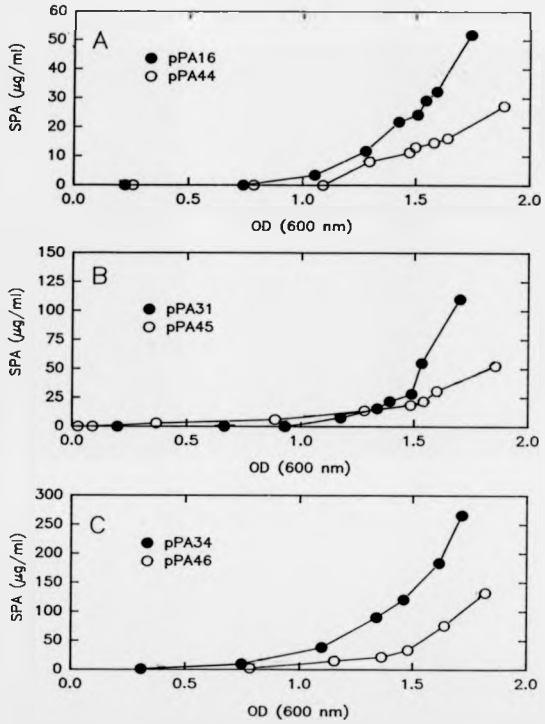


Figure 4.23 Comparative study to establish the effect of the vector–borne lac promoter upstream of the 5' terminus of the Spa gene

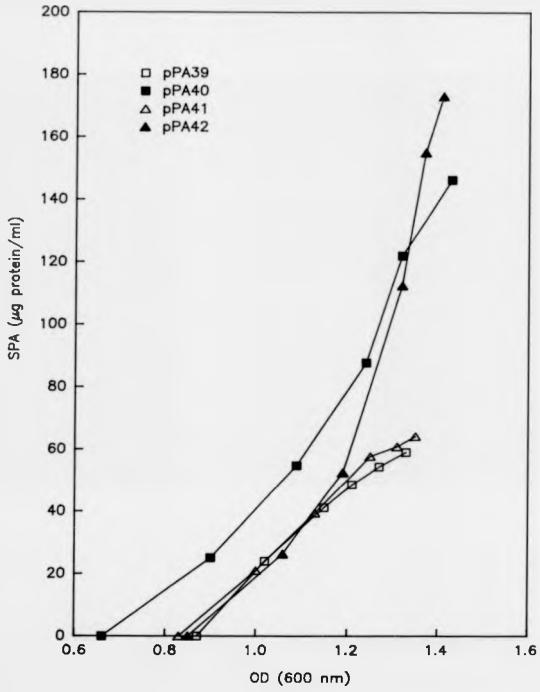


Figure 4.24 Effect on SPA production of: i alteration of the spa translational initiation codon [pPA39] & [pPA40] and ii. as i. plus removal of the spa promoter [pPA41] & [pPA42]

4.2.4 Fermentation of *E.coli* JM83[pPA] for the production of commercially viable amounts of SpA

4.2.4.1 Introduction

The cloning and manipulation of the *S.aureus* protein A gene detailed in this thesis has allowed the development of the SpA producing clones pPA31 and pPA34. When these are expressed in an *E.coli* host, produced under small-scale laboratory growth conditions, a 3-to 5-fold increase in the levels of SpA recovered is seen compared to the level of SpA production by *S.aureus* Cowan I. Unfortunately, these two clones have proven susceptible to plasmid loss and cell lysis unless grown under stringent conditions. Therefore, before these constructs can be considered a success in terms of commercial exploitation, it must first be shown that they can withstand the more vigorous conditions encountered during large-scale culture (150 L and 400 L) fermentation and subsequent downstream processing. It was decided to scale up fermentation of both JM83 (pPA34) and JM83 (pPA31) for two reasons. Firstly, the truncated SpA clone pPA34 produces a greater amount of SpA than pPA31 (2 fold increase), however, the latter clone has been shown to withstand more rigorous culture conditions. Secondly, the two *spa* recombinant clones produce SpA of different lengths, one with a hydrophobic cell wall binding C-terminus (pPA31) and one without this binding ability (pPA34), and so can be utilised for different purposes.

4.2.4.2 Experimental Procedures

The stability of the plasmid and viability of the host have been shown to be poor when maintained on antibiotic 2xYT or L-agar, with cell lysis occurring within a few days. Therefore, glycerinated cultures stored at -70^{o}C were used to seed broth to provide an inoculum for the 20 litre fermenter. Glycerinated seed cultures were prepared by growing cells in 50 ml of L-broth with 50 $\mu g\ ml^{-1}$ ampicillin in 150 ml shake flasks, at 160 rpm, 37^{o}C, for 8 hours. Ten ml of the culture was removed and tested for SpA levels while the rest was preserved with 15% sterile glycerol in 1 ml aliquots. The

aliquots were maintained for further use (providing the SpA assay showed an adequate level of SpA production). To prepare an inoculum for the 20 litre seed vessel, 50µl of a glycerol culture was inoculated into 10 ml of L-broth containing 50µg ml^{-1} ampicillin in a 25 ml universal, and shaken at 260 rpm, 37°C for 6 hours. This was then used to inoculate a litre of L-broth (50 µg ml^{-1} ampicillin) in a 2 litre unbaffled flask, which was shaken at 160 rpm o/n. Two litres of such culture was supplied to the CAMR Pilot Plant section to inoculate a 20 litre seed fermenter, this in turn was used to inoculate the main 150L or 400L culture vessel (when an OD_{650nm} of 2 was reached or after 6 hours incubation, whichever was sooner). The contents of the main culture vessel were harvested when the OD_{650nm} peaked (about an OD_{650nm} of 20).

Fermentation medium, and conditions for growth in the 20, 150 and 400 litre culture vessels are detailed in Appendix II.

During the fermentation procedures, 25 ml samples were taken and various parameters monitored. The pH and OD were monitored hourly and a cell purity and cell viability count performed every two hours. The SpA levels in the cells were also determined as follows. The monorocket SpA detection assay was performed on 10 ml volumes of culture, the cells were spun down and disrupted using lysozyme. The SpA assay was carried out on samples taken at the following points in the culture runs; on the 1 litre seed prior to inoculating the 20 litre vessel; on the 20 litre seed at hourly intervals from 4 hours growth and just prior to inoculating the main vessel; on the 150 litre and 400 litre main culture vessels at 4 hrs post inoculation and then at hourly intervals and just prior to harvesting and on a sample of the harvested paste. Ten ml culture samples from the 20 L, 150 L and 400 L culture vessels were spun down, the supernatant removed and reserved, the cells were then resuspended in 10 ml of growth medium and both supernatant and resuspended cells stored frozen for 48 hours until assayed. The growth medium in which the cells were resuspended (cell wash) was also assayed to determine how much SpA had been released from the cells during the freeze-thaw process.

Preparation and maintenance of the seed stock; the growth of the 10 ml and 1 litre seed cultures; and quantification of SpA levels (using the monorocket assay described in section 2.2.14.2) in culture samples from the seed and main culture vessels and in the final harvested cell paste were undertaken by myself.

Fermentation and monitoring of the culture in the 20 litre seed vessel, 150L or 400 L main culture vessels and harvesting of the cells were undertaken by shift technicians at the fermentation Pilot Plant under the guidance of Dr. S. Chambers, recovery and purification of the SpA to commercial standard was carried out by Dr. P. Hammond of the Diagnostic Reagents Group.

4.2.4.3 Results

Both JM83 [pPA31] and JM83 [pPA34] cultures were grown in identical 20 litre fermentation vessels, 10 ml culture samples taken at regular intervals, and the cells and supernatant assayed for SpA content. Results are presented in Table 4.13 and Figure 4.25. JM83 [pPA34] produces 17.46g SpA in a 20 litre culture and JM83 [pPA31] produces only 35% of this amount, 6.1g SpA in a 20 litre culture. Further examination of SpA levels, revealed that 76.5% of the SpA is cell associated in the JM83 [pPA34] culture whereas only 59% is cell associated in the JM83 [pPA31] culture. As it is only the cell paste that is harvested for recovery and purification of SpA for a commercial market, the use of JM83 [pPA31] would result in the discarding of 41% of the potential SpA yield, JM83 [pPA34] will limit the loss to 23.5% of that potential.

The freeze-thawing step involved in storage of the cells prior to assaying for SpA content was also seen to be highly detrimental to the cells, resulting in the release of SpA equivalent to 21.9% and 24.6% of that retained by pPA34 and pPA31 cell samples, respectively. Taking into account the above information JM83 [pPA34] was selected for large-scale fermentation in preference to JM83 [pPA31].

TABLE 4.13

Comparison of SpA Levels Obtained From a

20 Litre Fermentation Culture of JM83 [pPA31] and JM83 [pPA34]

JM83 [pPA34]				JM83 [pPA31]		
Incubation Time	SpA $\mu g\ ml^{-1}$ culture					
	lysed	Supn. cells	Whole culture	Lysed cells	Supn.	Whole culture+
1 litre seed culture-transfer	110	18	128	0	0	0
4hr	118	18	136	0	45	45
5hr	176	18	194	0	60	60
6hr	242	32	274	21	100	121
7hr	412	80	492	45	125	170
8hr	444	104	548	128	115	243
9hr	484	148	632	134	121	255
10hr	548	208	756	144	124	279
Total SpA *present	10.96 g	4.16 g		2.88 g	2.5 g	
%SpA in cell wash*	21.9%			24.6%		
*Potential SpA yield	13.3 g	4.16 g	17.46 g	3.59 g	2.5 g	6.09g

Cell wash*: Medium in which cells were resuspended and stored at -20°C following harvesting and removal of the supernatant (supn.). Assayed to ascertain the amount of SpA lost from the cells through damage caused by the freeze-thawing step used in storage. The final harvest of cells are stored as a frozen cell paste at -20°C, and thus have no wash step prior to the recovery of the SpA. The SpA $\mu g\ ml^{-1}$ level in the cell wash is presented as a % of the total SpA present in the lysed cells, as only every second hourly sample and the final sample were assayed for SpA (no % difference greater than 1.3 was observed). The SpA cell wash in the figure is presented as % as only every second sample and the final sample were assayed, none varied from the calculated average by more than 1.3%.

SpA whole culture+: SpA in lysed cells plus SpA in supernatant (excludes cell wash figure).

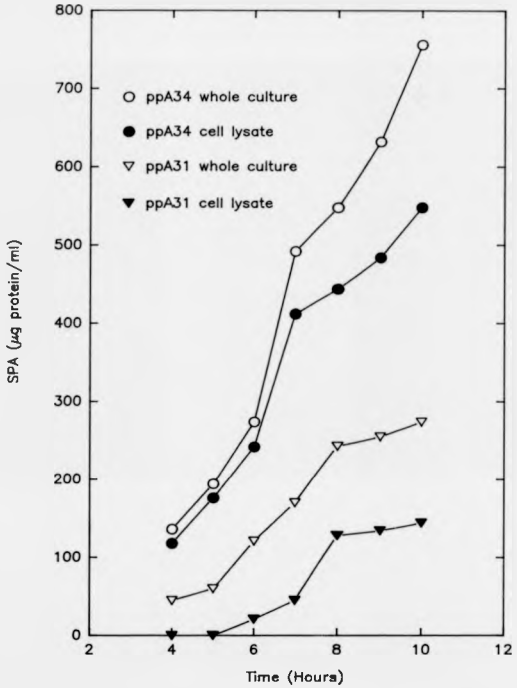


Figure 4.25 Graph comparing the amount of SPA present in intact cell lysates and in whole cultures during 10 hour, 20 litre fermentations of JM83[pPA31] and JM83[pPA34]

The samples of cell lysate and supernatant from JM83 [pPA34] were diluted to give a final concentration of 100μg ml^{-1} and were then examined, using the SDS gel fractionation and immunoblotting techniques described in section 4.1.2.3, to determine the size of the IgG binding polypeptides produced during the 20 litre fermentation, Figure 4.26. In the 4 hour cell lysate the predominant bands consist of the 48 kDa and 45 kDa SpA polypeptides, though less distinct, the 43 kDa and 32 kDa and 28 kDa polypeptide are also visible. There is a faint band seen in both the 4hr and 5hr cell lysates corresponding to 53 kDa in size, which is thought to represent the precursor form of the truncated protein coded for by pPA34. Degradation of the IgG binding polypeptides is seen to increase in direct proportion to the length of culture incubation, in particular, the 32 kDa and 28 kDa bands increase in intensity and there is also an increase in the amount of "non-banded" binding seen throughout the electrophoresis lane between the 48 kDa and 28 kDa bands. The supernatant samples show a very similar pattern of IgG binding polypeptides to that present in the cell lysates from the same time period, except that, overall, less degradation of the major polypeptide (48 kDa) was observed and the general increase in "background staining" within the lanes (probably caused by non-specific degradation) is less.

A large-scale fermentation run (400 litres) was carried out using JM83 [pPA34]. The results are shown in Table 4.14. The total culture yield was 195.7g, however 60% of this appeared to be present in the supernatant. Therefore, the amount recovered from the entire culture run was only 78.1g SpA in a total cell paste of 12,130g. Consequently, for a 20 fold increase in culture size, less than a 6 fold increase of recoverable SpA was seen. In the 20 litre seed culture, the amount of SpA in the cells was always considerably more than in the supernatant (Figure 4.27). Whereas, in the 400 litre main culture there was always more SpA accumulating in the supernatant than present in the cells (Figure 4.28). As SpA production appeared to decline following transfer from the 20 litre to the 400 litre container it was decided to try and grow JM83 [pPA34] in a smaller culture vessel (150 L) which did not have a 20 litre transfer stage.

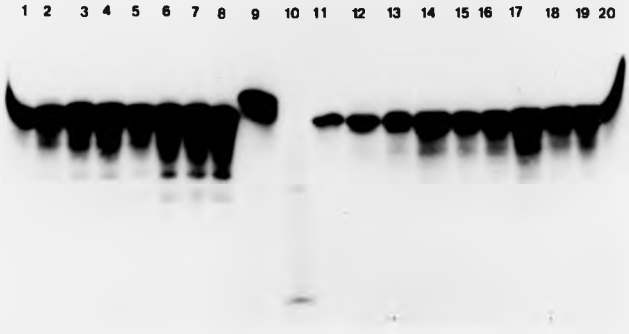


Figure 4.26 Analysis of IgG-binding Polypeptides Produced by JM83 [pPA34] Under 20 Litre Culture Conditions

All sample JM83 [pPA34] unless stated otherwise. Lanes contained

(1)	1 Litre seed culture cell lysate	(2)	20 cell culture 4 hr
(3)	20 L, cell culture 5 hr	(4)	20 cell culture 5 hr
(5)	20 L, cell culture 7 hr	(6)	20 L, cell culture 8 hr
(7)	20 L, cell culture 9 hr	(8)	20 L, cell culture 10 hr
(9)	*S. aureus* Cowan I	(10)	Molecular weight marker
(11)	1 litre seed culture, supernatant	(12)	20 L, supernatant 4 hr
(13)	20 L, supernatant 5 hr	(14)	20 L, supernatant 6 hr
(15)	20 L, supernatant 7 hr	(16)	20 L, supernatant 8 hr
(17)	20 L, supernatant 9 hr	(18)	20 L, supernatant 10 hr
(19)	20 L, cell paste supernatant, 10 hr	(20)	washed cells, 10 hr.

TABLE 4.14

SpA Levels Obtained Using a 20 Litre Seed Culture Vessel and a 400 Litre Main Culture Vessel for the Fermentation of JM83 [pPA34]

Culture Vessel	Time (hr)	OD_{600}	SpA µg ml^{-1} culture		
			Lysed cells	Supernatant	Whole culture
20 litre seed	5	2.78	33	25	58
	6	3.71	157	52	209
	7	5.50	171	59	228
	8	5.59	274	71	345
	9	7.09	368	96	464
	10	7.20	508	128	636
	Transfer		572	203	775
Total SpA present:			11.44g	4.06g	
% SpA in cell wash:			21.2%		
Potential SpA yield:			13.77g	4.06g	17.83g
400L main culture	4	3.08	44	86	130
	5	3.58	51	99	160
	6	6.42	100	120	220
	7	7.85	136	146	282
	8	12.75	149	164	313
	9	19.95	160	206	366
	9.5	23.25	173	294	467
Total SpA present:			69.9g	117.6g	
% SpA in cell wash:*			11.7%		
Potential SpA yield:			78.1g	117.6g	195.7g

+ and * see Table 4.13

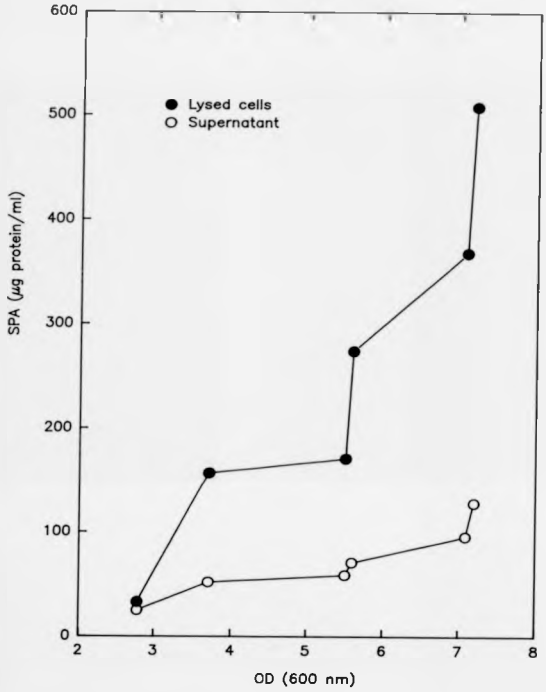


Figure 4.27 SPA production levels obtained using a 20–litre culture vessel for the growth of JM83[pPA34] seed culture for inoculation of a 400–litre culture vessel

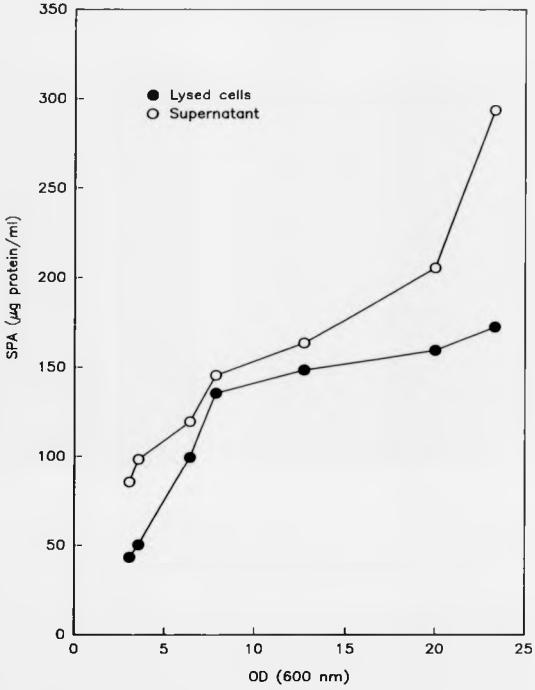


Figure 4.28 SPA production levels obtained using a 400–litre culture vessel for the growth of JM83[pPA34]

JM83 [pPA34] was grown in a 150 litre culture vessel for 13hrs (rather than the usual 10hrs) until an OD_{600nm} of 7.52 was reached (an OD_{600nm} of 7-8 gave optimum yields in the 20 litre cultures). Samples for SpA analysis were taken hourly from the 5hr point and the results shown in Table 4.15. A total culture yield of 167g was observed with 60% (101g) of the SpA associated with the cellular fraction. The total yield SpA/litre of the 150 litre culture is higher than that seen for both the 400 litre and 20 litre culture, with yields being 1.1g/litre, 0.49g/litre and 0.87g/litre, respectively. The SpA yields from the cellular fraction of the 150 litre, 400 litre and 20 litre were 0.67g/l, 0.2g/l and 0.67g/l, respectively. The amount of SpA in the supernatant can be seen to increase dramatically at the 12hr sample, OD_{600} 7.28. In the 11hr sample (OD_{600} 7.08) 81% of the total SpA was present in the cell (495µg ml^{-1} cell lysate (62.6%) and 18.4% more in the cell wash) whereas the 12hr sample (OD_{600} 7.28) the amount of SpA in the cell lysate was seen to fall, while the amount of SpA in the cell supernatant increased dramatically; so that only 64% of the total culture SpA is present in the cells. The dramatic increase in SpA in the supernatant was seen to continue in the 13hr sample (Figure 4.29).

Viable counts on cells from the 150 litre culture were performed using L-agar plates and L-agar ampicillin (50µgml^{-1}) plates. A comparison of the results allowed the percentage of the population with ampicillin resistance to be determined (Figure 4.30). As the coding capacity for ampicillin resistance and SpA production are carried on the same plasmid, the degree of ampicillin resistance of the culture will give a fair indication of the percentage of cells still capable of producing SpA.

TABLE 4.15

SpA Levels Obtained Using a 150 Litre Culture Vessel for the Fermentation of JM83 [pPA34]

Incubation Time (hr)	OD_{600}	SpA μg ml^{-1} culture		
		Lysed cells	Supernatant cells	Whole[+] culture
5	1.66	28	0	28
6	2.39	77	0	77
7	3.05	138	14	152
8	4.82	302	36	338
9	5.52	387	60	447
10	6.4	437	110	547
11	7.08	498	136	634
12	7.28	491	328	819
13	7.52	564	440	1004
Total SpA present		84.6g	66g	
% SpA in cell wash*		19.4%		
Potential SpA yield		101g	66g	167g

[+] and * see Table 4.13

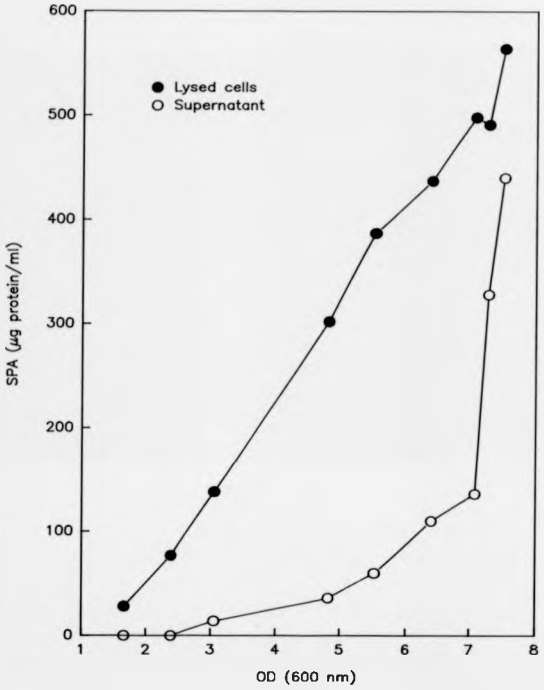


Figure 4.29 SPA production levels obtained using a 150–litre culture vessel for the growth of JM83[pPA34]

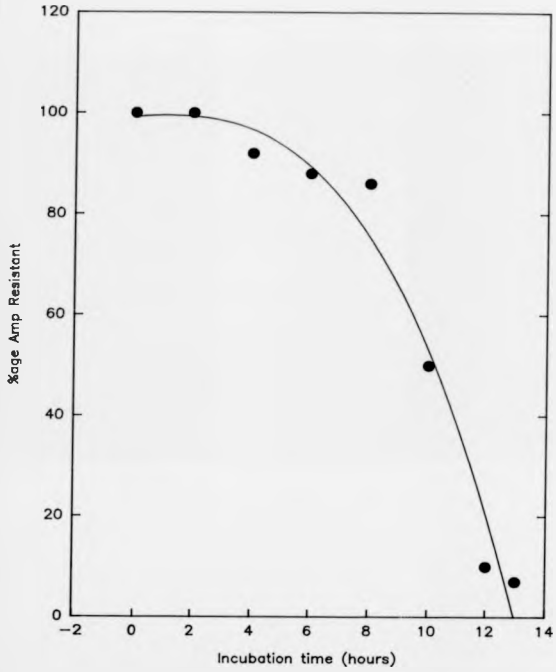


Figure 4.30 Graph showing percentage of population resistant to Ampicilllin (50 μg/ml)

4.3 DISCUSSION

Protein A has been successfully expressed from plasmids carrying the *spa* gene in *E. coli* (Duggleby and Jones, 1983), the level achieved, however, was no higher than that attained using *S. aureus* Cowan I. The work discussed here details the assessment of SpA producing plasmids, which were designed to produce higher levels of SpA in *E. coli* than those occurring naturally in *S. aureus* Cowan I. This assessment has been made in terms of the levels of SpA production achieved both under small-scale laboratory culture conditions and large-scale "commercial" culture conditions; and the effect of these levels of SpA production on the host organism's characteristics. Cellular localisation of the recombinant IgG-binding polypeptides have been examined in respect of the *spa* gene fragment utilised and expression levels in the cloned system have been compared with the naturally occurring SpA. Factors effecting SpA expression within the plasmid host system utilised have been investigated, and the results of each individual experiment discussed in Section 4.2; this section summarises those discussions and draws attention to the overall conclusions that can be made.

4.3.1 Background

Earlier work on SpA expression in *E. coli* carried out in this laboratory by Duggleby and Jones had utilised plasmid constructs which carried a truncated *spa* fragment, lacking the final 69 codons of the C-terminus. These constructs also carried 608 nucleotides of *S. aureus* DNA upstream from the *spa* gene. Duggleby and Jones (1983) indicated that the 608 bp region of *S. aureus* DNA, positioned between the N-terminus of the *spa* structural gene and the *lacZ* promoter of the pUC8 host plasmid might prevent transcription from the *lacZ* promoter having a discernible effect on *spa* expression. This suggested the possibility that high levels of *spa* expression might be achieved using the multicopy plasmid pUC8, if transcription from its *lacZ* promoter could be ensured. Using this data, three preliminary SpA clones (pPA30, pPA31 and pPA34) were constructed, and expression levels compared with those for pPA16. Comparison of the levels of SpA produced in overnight (10 ml cultures) by these four constructs

indicated that, not only did the removal of 608 bp between the plasmid *lacZ* promoter and the *spa* gene enhance the amount of SpA produced, there was also a 2-fold difference in SpA levels between those clones carrying the entire *spa* gene and those carrying the C-terminally truncated *spa* gene. Possible reasons for these differences in SpA production will be discussed later (4.3.3).

4.3.2 Protein A Production in *E. coli*

Comparison of the SpA levels produced by the four *E. coli* clones with that of the native *S. aureus* Cowan I, showed that both pPA31 and pPA34 produced substantially more: 2-3 fold higher and 4-5 fold higher, respectively. Therefore pPA16 producing an equal amount of SpA to *S. aureus* Cowan I, pPA31 and pPA34 were considered as potentially suitable for commercial exploitation and were subject to further investigations. pPA30 was included in many of these studies to allow a full comparison to be made.

Analysis of the IgG binding peptides produced by the SpA plasmids demonstrated a major difference between the SpA produced by *S. aureus* Cowan I and the *E. coli* recombinant clones. This was revealed by analysis of the IgG binding peptides isolated from samples derived after culturing the organisms for 16hrs. Whereas *S. aureus* Cowan I contained only one form of SpA, the *E. coli* clones were shown to contain up to 7 different sizes of IgG binding protein. The largest of these was found in clones carrying plasmids encoding the full length SpA and were only 2 kDa smaller than the *S. aureus* Cowan I SpA. The largest polypeptide from the plasmid encoding the truncated SpA is 7 kDa smaller than the Cowan I Spa, the size expected allowing for the lack of 69 a.a at the C-terminus. The rest of the small size IgG-binding polypeptides are thought to represent degraded forms of the largest polypeptide. The degradation appears to result from site specific proteolytic cleavage, rather than C-terminal "chew back", as discrete bands, rather than a general background smear are seen in the electrophoresis lanes and polypeptides of identical molecular weights are present in the lysates from the majority of clones. These peptide cleavage points do not appear to be

located in the N-terminal, IgG-binding portion of the SpA molecule as N-terminal sequence analysis of the IgG-binding polypeptides produced by JM83 [pPA34] have shown all the polypeptides to possess the same N-terminal sequence; that of native SpA. Degradation is thought unlikely to be the result of *E. coli* serine protease activity, as the inclusion of the general serine protease inhibitor (phenylmethyl-sulphonyl fluoride) did not alter the binding pattern observed. However, not all serine proteases are sensitive to this reagent. Time course experiments showed that the levels of the various size IgG binding proteins varied at different times in the culture; in general, the larger IgG binding polypeptides was seen first with the smaller 'degraded' forms appearing with increasing time after inoculation. Again, this indicated that these were degradation products of the larger forms. The full length SpA appeared to be resistant to degradation for longer than the truncated SpA. Towards the end of the culture run, about 16hrs post inoculation, the largest IgG-binding polypeptide form was seen to be absent from the cell lysates of JM83 [pPA34] whilst the level of the smaller IgG binding peptides remained constant, or was seen to increase further, this also corresponds to a decrease in the total amount of SpA in the cell lysate and an increase in the amount of SpA in the supernatant. This would indicate that *de novo* syntheses of SpA was decreasing prior to cell lysis when a culture was grown under optimum conditions. The strong correlation between the sudden decrease in cell lysate SpA levels with a sudden increase in SpA in the cell supernatant; would suggest that SpA is not actively secreted to an extracellular location but is probably in the supernatant as the result of cell lysis or increased cell permeability.

Scale up of the culture volumes involved in the time course experiments from 10 ml to 500 mls, revealed a problem of cell fragility that had not been apparent in the small-scale samples. JM83 [pPA31] and [pPA34] which produced the highest levels of SpA were found to be extremely susceptible to cell lysis when subjected to stress. For instance, culture aeration had to be low, as anything more than modest agitation would cause lysis, and growth on a minimal media would result in early lysis. The use of alternative *E. coli* hosts for expression of the two high SpA producing plasmids, pPA31

and pPA34, did not improve viability of the cells during culture or storage and the SpA yields between the different strains did not vary greatly from that of JM83. Scale up to Pilot Plant production levels of 20 litre, 150 litre and 400 litre cultures showed JM83 [pPA31] to be move liable to cell lysis than JM83 [pPA34]. As the established practice for recovery of SpA from cell cultures involved only the harvesting of the cells, cell lysis prior to harvesting resulted in a significant loss of SpA. Therefore, it was important to prevent cell lysis prior to harvesting the large-scale cultures. JM83 [pPA34] was successfully grown to produce large amounts of SpA in both the 20 litre and 150 litre fermentation vessels, with 78.1% from a total yield of 17.46 g SpA, and 60% of 167 g SpA, being retained in the cells, respectively. JM83 [pPA34] produced 195.7 g of SpA when grown in the 400 litre fermentation vessel, however only 40% of this was present in the cells, the remainder was in the supernatant. Thus, the overall amount of SpA recovered from the 400 litre fermentation run was less than from the 150 litre fermentation run. It was noted that the percentage of SpA found in the cell wash for the 400 litre culture samples was approximately half that found in the 20 litre and 150 litre cultures. This seems to indicate that the surviving cells are less fragile, which suggests that a possible reason for the low amounts of SpA produced in the whole culture may be a result of constant level of cell lysis in cells containing a high level of SpA. This would also explain the very high levels of the SpA in the supernatant compared with the lysed cells. It was also observed that SpA production appears to be at its maximum when the culture has an OD_{600} of 7; after which an increase in cell lysis is observed.

It would appear that two separate factors are responsible for lysis. Firstly, the levels of SpA in the cell and secondly, the physical agitation the cell receives during growth. Under the gentle aeration conditions used in the small-scale cultures, JM83 [pPA34], the highest SpA producing plasmid, will lyse when the culture medium becomes depleted of nutrients. Under the more vigorous growth conditions of a large-scale culture, JM83 [pPA31] will lyse; even though cell growth is still exponential. This

suggests that the possession of the hydrophobic membrane region of the SpA molecule is responsible for the fragility of JM83 [pPA31]. It is proposed that this region of the SpA molecule may bind to the membranes in *E. coli*, and, in doing so, interfere with the sites normally concerned with septate formation and cell division. This is supported by the morphological appearance of cells examined using light microscopy, where JM83 [pPA31] was seen to be up to 20 times the normal length, this extreme elongation of cells has often been reported to occur when the high levels of heterologous protein are produced (Schoner *et al.*, 1985; Schoemaker *et al.*, 1985; Georgiou *et al.*, 1986). Although JM83 [pPA16] and JM83 [pPA34] were the same length as non-plasmid carrying JM83, they both appeared slightly more ovoid in shape than the normal JM83 rod. This would appear to indicate that cellular elongation is a genuine result of the hydrophobic C-terminus rather than an effect of over-expression and the increase in the cellular SpA levels. The thesis is also supported by the fact that the detectable cellular SpA levels in JM83 [pPA31] increased when the cells were disrupted by sonication, rather than lysozyme/detergent lysis where there is a greater chance of membrane associated SpA being trapped by cell debris, pelleted and discarded. If the SpA produced in JM83 [pPA31] is more likely to be membrane associated than the SpA produced in JM83 [pPA34], this may also mean that the some of the apparent 2-fold difference in SpA levels first indicated may be a difference in the rate of recovery, rather that the rate of production.

Cellular localisation of the SpA to the different compartments of the cell using traditional fractionation techniques proved unsuccessful using JM83 [pPA31] and JM83 [pPA34] as the degree of cell fractionation was poor, with contamination occurring between the cytoplasmic, periplasmic and membrane fractions. This was probably due to the fragile nature of both high SpA producing clones; particularly under phosphate limited conditions. In addition, both these clones failed to produce measurable levels of SpA when grown for 6 hours on a low phosphate broth. However, the JM83 [SpA] clones producing lower levels of SpA, pPA16 and pPA30 were successfully fractionated using the above method and also produced a reasonable level of SpA.

Both the full length SpA and the truncated SpA were found to be located mainly in the periplasm (78.8% and 81%, respectively), and the full length carrying the hydrophobic "membrane-binding" C-terminus was found in the membrane fraction at a level (7.6%) only slightly higher than the truncated SpA form (5.4%. Further examination of the cell fractions of JM83 [pPA30] to allow visualisation of the different molecular weight IgG binding peptides indicated that most of the proteolysis of SpA occurred in the periplasm, with very little full size SpA being present. Most of the SpA associated with the cell membranes was full size Both full size and degraded IgG binding polypeptides seen in the cytoplasmic fraction can be accounted for by the levels of contamination seen in the localisation assay results (Table 4.5). A less disruptive method of localising the SpA within the cell, using immuno-gold labelling and electron microscopy, was also attempted.

Electron lucent areas were seen in *E. coli* expressing SpA. These areas could be seen to be present between the inner (cytoplasmic membrane) and the outer cell membrane (Figure 4.10 and 4.11), thus occupying the periplasmic space. They appear to increase the size of the periplasmic space considerably, as normally both membranes are seen to be in close opposition. The aim of trying this method was to assess its suitability for determining the location of SpA within the cell by probing cross sections of JM83 [pPA16] with immunogold. Results showed the immunogold to be present almost exclusively in the electron lucent areas which would indicate that the truncated SpA protein coded for by pPA16 is transported to the periplasmic space. This result concurs with that reported using the traditional localisation studies (Section 4.2.1.9 and 4.2.1.10). There appears to be no significant association with either the inner or outer membrane of the cell, although some gold particles are seen along the outer membrane and cell wall (Figure 4.13). This may be an artifact of cross sectioning of the cells. Whilst the electron lucent areas may represent accumulations of SpA, they could also represent an accumulation of β-lactamase, which has been shown to be highly expressed by the [SpA] progenitor plasmid [pUC8] (unaltered *bla* gene) under identical

growth conditions to those used for JM83 [pPA16] (Warnes 1989). Georgiou *et al* (1986) also reported β-lactamase to be localised in the form of electron dense inclusion bodies present in the periplasm rather than in the cytoplasm. Further experiments involving electron microscopy of *E. coli* JM83 and *E. coli* JM83 [pUC8] would help resolve this problem. The negative results of Warnes for immunogold binding to SpA *E. coli* lacking the SpA plasmid, and the presence of the gold label bound to the surface of *E. coli* JM83 [pPA16] shown in 4.12, indicates that some antigenic portion of a small number of the SpA molecules protrude through the outer cell membrane and wall. Preliminary studies on pPA16 would appear to indicate that immuno gold electron microscopy can be used to determine the localisation of SpA within an *E. coli* host at low levels of expression. Further experimentation on JM83 [pPA31] and JM83 [pPA34] using this method is necessary to determine if this method can be used successfully on the more fragile high expression SpA clones.

Subsequent studies by Warnes (A. Warnes personal communication) on JM83 [pPA31] have shown the SpA to be localised in the periplasmic space, both in the electron lucent areas and also around the outer membrane of the host. This may be due to the fact that pPA31 contains the hydrophobic membrane binding region of protein A which is not present in pPA16 or pPA34. Alternatively, it may be a result of higher levels of SpA present in the host cell carrying pPA31. Results from pPA34, which is identical to pPA31 except it lacks the hydrophobic C-terminal sequence, may explain this. If the hydrophobic C-terminus of pPA31 is binding to the membrane, this could well interfere with septate formation of the host cell and so prevent division. This would also explain the extremely elongated forms of JM83 associated with the presence of pPA31 but not pPA34 or pPA16. Analysis of the effect of the lower expression levels of full length SpA produced by pPA30 on the cell structure may well help resolve these results. However, the fact that it only produces a low level of SpA may well not interfere with septate formation. Warnes showed electron lucent areas to be present in JM83 [pUC8] and that the levels of β-*lactamase* activity found in the cell fell as the SpA levels increased. It is possible that both of these proteins may well compete for export to the

periplasm. Thus β-lactamase may not be a good choice of selective marker for a high heterologous protein production strain secreting the protein.

As cell fragility appears to be an inherent characteristic of high levels of SpA production in *E. coli*, there would appear to be two options available for commercial exploitation of the SpA clones i) alter the fermentation procedure so that the cells survive longer and alter the harvesting procedure to allow recovery of all the SpA produced; whether from intact cells, or the culture medium (or ii) alter the *E. coli* clone so that *spa* expression is under greater control and production of SpA can be limited to the point of the culture growth cycle where maximum SpA culture levels can be achieved.

4.3.3 Gene Factors Effecting *spa* Expression

This area of work was undertaken not only to try and control the production of SpA during a culture but also to investigate the efficiency of the Gram-positive *spa* promoter in a Gram-negative *E. coli* host and to establish factors which might effect this efficiency in *E. coli*.

Initial cloning of the *S. aureus* 2.012 bp fragment containing the *spa* gene into a multicopy (500 copies per cell) vector containing the strong *E. coli lac* promoter, appeared to show only a small increase in SpA production which could be accounted for solely by the increase in copy number. This suggested that either the *spa* gene could not be expressed using the *lac* promoter, or that transcriptional readthrough from the *lac* promoter was being prevented entirely or interrupted from reaching the *spa* gene. Examination of the nucleotide sequence of this region revealed one area of dyad symmetry of a reasonable length (Chapter 3, Section 3.3.2.3). This did not correspond to a transcriptional termination sequence. It is of interest to note the presence of two sequences within this region, that closely resemble the CRP - binding sites and other operator sites of *E. coli* (Ebright, 1982; de Crombrugghe *et al.*, 1984, Studnicka, 1987)

and the regulatory sites found in catabolite-repressed *B. subtilis* promoters (Weickert and Chambliss, 1990). Therefore, it was considered likely that this fragment of *S. aureus* might interact with the cAMP/CRP complex and so prevent readthrough of transcription from the *lac* promoter. Removal of this region did indeed result in a large increase in the levels of SpA produced; (a 12-fold increase between clones JM83 [pPA30] and JM83 [pPA31] and a 4-fold increase between clones JM83 [pPA16] and JM83 [pPA34].

Transformation of these plasmids into JM101, which was inducible for *lac* promoter expression and which could be induced using IPTG, showed that the *lac* promoter contributed to at least 17% of the SpA level seen in pPA31 and pPA34, but only contributed to 2.1% of the level seen in pPA16. This would be consistent with the proposed interference in transcription from the *lac* promoter. However, due to the well known "titration" effect seen when trying to control a multicopy plasmid promoter from a low copy expression system it was not possible to conclude whether the remaining SpA expression was due to the *spa* promoter activity or to *lac* promoter activity caused by a comparatively low amount of repression protein. Attempts to determine this by increasing the amount of repressor protein 10-fold, using the plasmid pNM52, were found to actually increase the amount of SpA when no inducer (IPTG) had been added to the growth medium. Levels of SpA increased still further when the *lac* promoter was derepressed. It would appear that either total repression of the *lac* promoter allowed increased expression from the *spa* promoter (as occurs with the CRP activated PII promoter of the *lac* operon in *E. coli*, where the CRP binding protein prevents the binding of RNA polymerase to the weak PI promoter (Munsen *et al.*, 1984; Malan and McClure, 1984); or, that the *lacI* repressor protein in someway binds to the sequence upstream of the SpA promoter and increases expression; any binding site for the *lacI* repressor protein must be located in the region 89 bp prior to a structural gene, as the same response was seen to occur in all three plasmids, and pPA34 and pPA31 only carry this 89 bp region of non-SpA *S. aureus* DNA, unlike pPA16 which carries 608 bp of *S. aureus* (including the two "CBS" homologous sequences) prior to *spa* standard

gene. Alternatively, pNM52 could have introduced an unknown factor into the system. It was clear that, whichever, if any, of the above might be happening, it was not possible to determine what level of *spa* expression could be attributed to each of the promoters whilst both were present on the same plasmid. Removal of the *spa* promoter had no discernible effect on the production of SpA, both in JM83 and JM101 (with or without IPTG), when compared to the levels of SpA of the corresponding *spa* promoter$^+$ plasmids, indicating that the *lac* promoter was capable of directing high expression of the *spa* gene. The *spa* promoter, however, was shown to be capable of directing only 50% of the *spa* expression compared with that seen using the *lac* promoter alone.

It was of interest to note that when the *spa* was removed from the transcriptional control of the *lac* promoter (thus effectively being expressed from the *spa* promoter alone), that the levels of SpA production for JM83 [pAP44] compared with those of JM83 [pPA46], although all reduced, retained the 4-fold difference previously seen between pPA16 and pPA34. As the result could no longer be attributed to interference with transcription from the *lac* promoter, and the only remaining difference between the two sets of clones was the presence of the 608 bp *S. aureus* DNA in the set (pPA16 and pPA44) producing lower levels of SpA, it would appear that this region must in some way decrease *spa* promoter efficiency. Examination of this region, as reported earlier, revealed several features that could possibly cause such an effect. Firstly, the presence of 2 regions with a high degree of homology to CRP binding site sequences and other reputed operator sites in *E. coli* (Studnicka, 1987) and to the proposed regulatory site of catabolite repressed *B. subtilis* promoters (Weickert and Chambliss, 1900). Binding of the CRP or another operator protein may produce a conformational change in the adjacent DNA lowering the efficiency of binding of RNA polymerase to the *spa* -35 and -10 sequences. The affect that CRP binding sites have, if any, on *spa* expression could be investigated by using *crp*$^-$ or *cya*$^-$ mutant hosts. Abrahmsen *et al.*, (1986) has suggested the presence of a stress-induced promoter located within this 608 bp region,

(section 3.3.2.3) which is functional in *E. coli*. This again may interfere with the efficiency of the binding of RNA polymerase to the *spa* -35 and -10 sequences located immediately prior to the *spa* gene. However, the results of Abrahmsen *et al.*, (1986) indicate an increase in SpA levels when this promoter is functioning, rather than a decrease. It is possible that the presence of an "CRP" binding site or operator sequence affects this promoter.

The *spa* gene is known to be subject to regulation by the *agr* locus and the associated δ-lysin gene (*hld*) in *S. aureus* (Recsei *et al.*, 1986). Disruption of the *hld* transcript has been shown to enhance the expression of *spa* (Janzon and Arvidson, 1990). It is possible that the 608 bp region contains a sequence that acts as a regulatory site for *spa*, which if "activated" decreases *spa* expression, removal of this region may well result in the increase in SpA levels observed. More studies are needed to establish the regulatory function of the DNA sequence in this 608 bp region, both in *E. coli* and in the natural host *S. aureus*.

CHAPTER 5

CLONING AND NUCLEOTIDE SEQUENCE ANALYSIS OF THE *E.COLI* (JM83) ALKALINE PHOSPHATASE GENE

5.1 INTRODUCTION

Alkaline phosphatase (orthophosphoric-monoester phosphohydrolase, EC 3.1.3.1) is a metalloenzyme specified by a single gene (*phoA*) in *E. coli* (Berg, 1981). The active form of the enzyme is secreted through the cytoplasmic membrane to its final location in the periplasmic space where it exists as a dimer of two identical 450 amino acid polypeptide chains. This secretory process involves the use of a signal peptide, as alkaline phosphatase is initially synthesised as a larger polypeptide chain with a mainly hydrophobic N-terminus which is cleaved after entry of the protein into the periplasm (Chang *et al.*, 1980; Michaelis and Beckwith, 1982). Alkaline phosphatase occurs in three active forms. These forms, or isoenzymes, differ only in the presence of an N-terminal arginine residue on both subunits of isoenzyme I, the absence of such a residue on both subunits of isoenzyme III, and in isoenzyme II by the possession of an arginine residue on one subunit but not on the other (Kelley *et al.*, 1973; Schlesinger, 1975).

The expression of the *phoA* gene is controlled by a complex regulatory mechanism consisting of a series of non-contiguous genes, often referred to as the *pho* regulon, this includes both positive (*phoR*, *phoB* and *phoM*) and negative (*phoR*) regulatory genes (Wanner and Latterell, 1980; Tommassen and Lugtenberg, 1982). Transcription from the *phoA* gene is stimulated in response to the amount of phosphate available to the cell. When low levels of phosphate in the medium become the limiting factor of growth, the *phoA* gene is totally derepressed and results in alkaline phosphatase production to levels representing up to 6% of the total cell protein (Garen and Levinthal, 1960). Under conditions of high phosphate the amount of alkaline phosphatase synthesised is negligible.

The precise physiological function of alkaline phosphatase in *E. coli* is not clear. The enzyme is non-specific and hydrolyses all phosphates at similar rates, regardless of their size or chemical nature. Therefore, its role may be solely to supply phosphate from phosphate esters under conditions of phosphate deprivation. However, production

of alkaline phosphatase in the majority of bacteria is constitutive, rather than inducible. This has led Reid and Wilson (1971) to suggest a possible additional role for the enzyme in phosphate transport, arguing that the provision of phosphate under conditions of deprivation is unlikely to be the enzymes only purpose.

Alkaline phosphatase has long been used as a marker enzyme due to the availability of a very simple and rapid colourimetric assay (p-nitrophenyl-phosphate is used as a substrate and the enzyme hydrolyses the bound phosphate to liberate yellow p-nitrophenolate). The main problem in the preparation of enzyme-labelled antigens or antibodies, for use in the detection and quantification of biological substances, is the reduction in enzyme activity often observed after the chemical procedures used to produce the conjugate. The construction of a *spa::phoA* gene fusion will result in the production of a defined bi-functional hybrid protein which has both the full enzymatic activity of alkaline phosphatase and the IgG binding ability of SpA. Such a bifunctional hybrid protein can be put to a wide range of immunological uses, as discussed in detail in section 1.1.5.

The production of a well defined gene fusion is considerably facilitated by the availability of the nucleotide sequences of the genes involved. These data allow the two genes to be precisely fused, translationally in phase, using either naturally occurring restriction sites, or sites created at the desired position(s) using site-directed mutagenesis. However, although the complete amino acid sequence of alkaline phosphatase isoenzyme III has been determined (Bradshaw *et al.*, 1981), only a small portion of the *phoA* gene sequence is known. The sequence determined is from the 5′ -end of the gene, and equates to the promoter region and codons specifying the signal peptide and first 56 amino acids of the mature protein (Kikuchi *et al.*, 1981; Inouye *et al.*, 1982). To facilitate future fusion of the *phoA* gene to *spa*, the gene was cloned from the *E. coli* strain JM83, and its entire nucleotide sequence determined.

5.2 EXPERIMENTAL AND RESULTS

5.2.1 Detection of an active *phoA* gene in *E. coli* JM83

The *E. coli* K12 strain chosen as the source of the *phoA* gene was strain JM83, a bacterial host in common use in this laboratory. Before commencing cloning experiments it was first necessary to establish the presence of a functional *phoA* gene. The nucleotide sequence of the 5′ end of the *E. coli phoA* gene has been determined by three independent research groups (Boidol *et al.*, 1982; Inouye *et al.*, 1982; Kikuchi *et al*, 1981). All report an identical nucleotide sequence for the region from the Pribnow box (-10 region) to the 27th codon of the structural gene. This information was used to design a 17 base oligonucleotide probe which could be used to detect the *phoA* gene by DNA/DNA hybridisation methodologies (ie., Grunstein-Hogness colony hybridisations and Southern blots). The probe (5′-CTATTGCACTGGCACTC-3′) corresponds to a region in the signal sequence 14 bp to 30 bp from the first nucleotide of the GTG initiation codon.

E. coli JM83 chromosomal DNA was cleaved with various restriction enzymes and the restriction fragments obtained transferred, following agarose gel electrophoresis, onto a nitrocellulose membrane. The nitrocellulose was then incubated with the ^{32}P-labelled oligonucleotide probe at 45^{o}C, using the conditions and buffers described in Section 2.2.9. Consistent with the restriction map of the previously cloned *phoA* gene (Inouye *et al.*, 1981), a 2.7 kb *Xho*I/*Hin*dIII fragment of the JM83 chromosome was shown to hybridise strongly to the oligonucleotide probe. In parallel, *E. coli* JM83 and *E. coli* AW1046 (a negative control, lacking the *phoA* gene) were grown overnight on low phosphate (LP) indicator plates (see Section 2.1.4) and in LP broth. The latter was then subcultured and the cells grown until exponential phase and then assayed (see section 2.2.18.2). The resulting colour change observed in assays of *E. coli* JM83 cell lysates were indicative of alkaline phosphatase production, whereas no colour change was observed using cell lysates derived from *E. coli* AW1046.

5.2.2 Construction of an *E. coli* JM83 gene bank

Because the source of the gene being cloned was the same bacterial species as the intended recombinant host, it was considered prudent to employ a *recA* derivative, JLM85. The use of such a host should obviate any problems of recombination between cloned plasmid inserts and chromosomal DNA. A total of 450 µg of *E. coli* JM83 chromosomal DNA (prepared as detailed in Section 2.2.4.1) was partially digested with *Sau*3A (see Section 2.2.3.8), the restricted DNA heat inactivated by a 10 min incubation at 70°C, and the restriction fragments produced size fractionated by electrophoresis on a 1% agarose gel. DNA fragments of between 10 kb --> 15 kb were recovered from the gel by electroelution (see section 2.2.4.6), ligated into *Bam*HI-cleaved pUC8 and transformed *E. coli* JLM85 (rubidium chloride method, Section 2.2.2). Transformed cells were plated onto L-agar ampicillin (50 µg ml^{-1}) BCIG indicator plates and incubated overnight.

Presumptive recombinant clones (colourless colonies) were replica picked onto a L-agar ampicillin 50 (µg ml^{-1}) master plate and onto a nitrocellulose grid placed on an surface of the L-agar ampicillin (50 µg ml^{-1}) plate, and incubated overnight at 37°C. Each clone was also individually transferred into a microtitre tray well containing 100 µl of L-broth, ampicillin (50 µg ml^{-1}). The microtitre trays were incubated overnight at 37°C and then glycerol was added to each well, at a final concentration of 20%. The trays were then frozen at -70°C, forming an *E. coli* JM83 gene bank for future use.

5.2.3 Detection of the *phoA* gene from an *E. coli* gene bank

Clones carrying the *phoA* gene were detected by *in situ* colony hybridisation using the ^{32}P-labelled oligonucleotide as a probe. Hybridisation of the nitrocellulose filters was undertaken by the procedure of Grunstein and Hogness (1975), at a temperature of 45°C. (see Section 2.2.7.2 for details). Following autoradiography of the processed filters, 7 presumptive positive clones were detected.

All the putative positive clones were restreaked and their plasmid DNA extracted using the STET small-scale isolation procedure (Section 2.2.4.2.1). The DNA was cleaved with the restriction enzyme combinations *Hin*dIII/*Xho*I and *Eco*RI/*Pst*I, and the restriction fragments generated analysed by agarose gel electrophoresis. Although all the recombinant plasmids released the expected pUC8-derived 2.7 kb fragment following digestion with *Eco*RI/*Pst*I, differences in the size and number of additional fragments suggested that no one clone was identical to another. Of the 7 clones, only two (2/47 and 9/27) released the expected 2.7 kb *Hin*dIII/*Xho*I fragment (Figure 5.1a). Two other clones had a slightly smaller band of 2.4 kb present in the *Hin*dIII/*Xho*I track. The DNA fragments were transferred to a genescreen membrane, and hybridised with the ^{32}P-labelled oligonucleotide probe, using identical conditions to that described earlier. The 2.7 kb *Xho*I/*Hin*dIII fragments of clones 2/47 and 9/27 were shown to hybridise strongly to the probe (Figure 5.1b).

5.2.4 Sub-cloning and Restriction Mapping of the *phoA* Gene

A large scale plasmid preparation was performed on clone 2/47, and the resultant plasmid cleaved with *Hin*dIII/*Xho*I and the 2.7kb fragment isolated from a 1% agarose Tris/acetate gel by electroelution. This fragment was then ligated into the complementary *Hin*dIII/*Sal*I sites of pUC8 and the resulting clone transformed into competent JLM85. Thereafter, a restriction map of the *phoA* encoding insert was generated by comparison of the DNA fragment sizes obtained following digestion with various restriction enzymes, either singularly or in combination. These included; *Bam*HI, *Pvu*II, *Eco*RI, *Bam*HI/*Pvu*II, *Bam*HI/*Eco*RI, *Eco*RI/*Hin*dIII and *Eco*RI/*Pvu*II (Figure 5.2).

5.2.5 Site-Directed Sequencing of the *phoA* Gene

DNA from clone 2/47 was cleaved either with *Pvu*II, or *Hin*dIII and *Eco*RI, and electrophoresed on a 1% agarose gel. The 1.96 kb, 0.78 kb and 0.21 kb *Pvu*II

Figure 5.1 **Detection of a Positive Alkaline Phosphatase Clone from a JM83 Gene Bank in J1.MR5**

A. 1% Tris/acetate gel showing restriction fragments produced for positive *pho*A clones: (a) 9/43, (b) 9/27, (d) 5/16, and (e) 2/47. Tracks (from right to left) 1–3 of each digested clone represents; (1) uncut (2) *Eco*RI/*Pst*I and 3. *Hin*dIII/*Xho*I. m = 2.7 kb pUC8 marker. The putative 2.7 kb alkaline phosphatase containing fragment is arrowed.

B. Autoradiograph of gel shown in Figure 5.1 A, after probing for the presence of the *pho*A gene. Restriction enzyme digests are: U = uncut, E/P = *Eco*RI/*Pst*I and X/H = *Xho*I/*Hin*dIII.

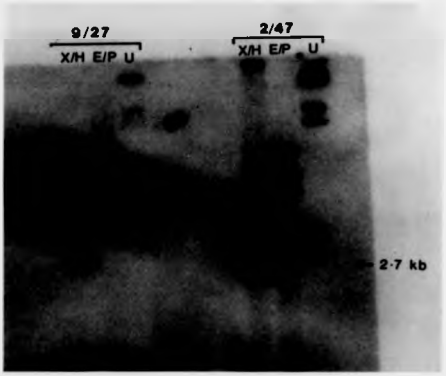


Figure 5.2 Restriction Map of the 2.7 kb *HindIII-XhoI* Fragment Containing the *PhoA* Gene of *E. coli* JM83

(A)

The restriction fragments used for site directed sequencing are indicated along with the amount of sequence data generated (200 - 300 bp) from each fragment (⟼).

The position of sequences complementary to the oligonucleotide primers are indicated △ , and their direction of synthesis (⟼). The sequence of the oligonucleotide primers were:

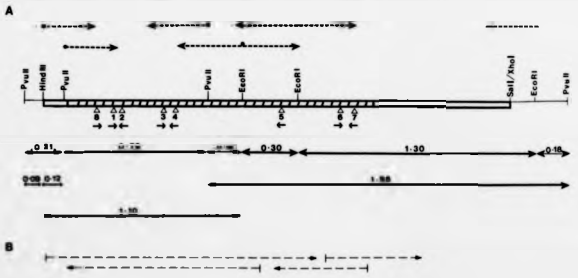
1. CAC CCG GCG GTG CTC
2. TAA GAG AAT CAC GCA
3. TCA GCA ACC GGC TGG
4. CCC AGC GCG CCG TTA
5. TCC TGC AAG TTG AAG
6. ACT GAC CGA CCA GAC
7. GGG CGC GGT TTT ATT
8. CTA TTG CAC TGG CAC TC (also used as the alkaline phosphatase gene probe)

(B)

Map of the alkaline phosphatase showing the regions of the gene covered by dideoxy-sequencing

5⟼3 strand sequencing ⟼ 86% nucleotide sequenced
3⟼5 strand sequencing ⟼ 89% nucleotide sequenced

all regions indicated ⟼ represented the consensus sequence derived from the sequencing of at least 4 different templates for that region.

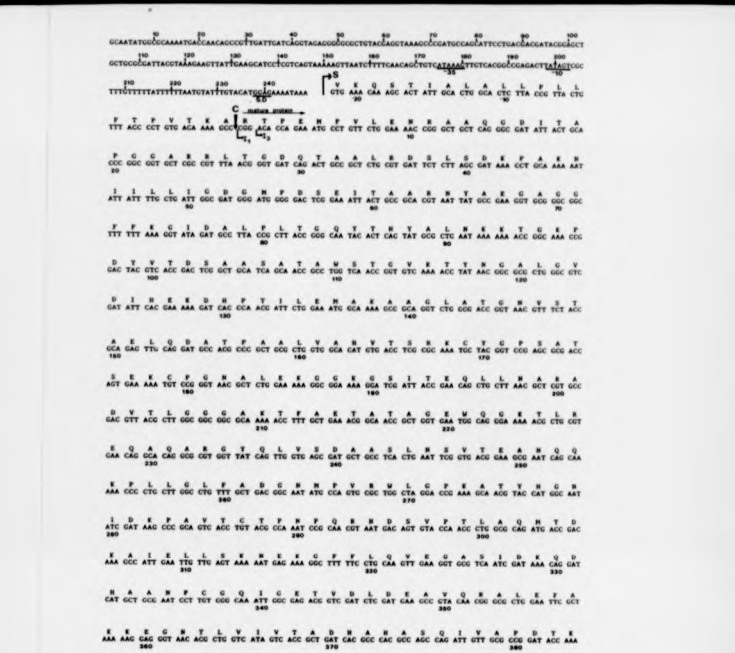


fragments, the 0.3 kb and 1.3 kb *Eco*RI fragment, and the 1.1 kb *Hin*dIII/*Eco*RI fragment (Figure 5.2) were gel purified by electroelution and cloned into appropriately cleaved M13 vectors. In the case of the three *Pvu*II fragments, the vector employed was *Sma*I cleaved, dephosphorylated M13mp18 DNA. The isolated *Eco*RI fragment was cloned into *Eco*RI cleaved and dephosphorylated M13mp18 DNA, while the 1.1 kb *Hin*dIII/*Eco*RI fragment was inserted into dephosphorylated M13mp18 and mp19 DNA, cut with *Hin*dIII/*Eco*RI. The ligation mixes were transformed into competent *E. coli* TG2, plated out onto L-agar and covered with H-top agar containing BCIG.

Five positive (white) plaques were picked from each of the above transformations, template DNA prepared and their inserts sequenced (as described in Section 2.2.10.5) using universal primer. Thereafter, the sequence information obtained from individual templates was extended by using a sequence-specific, custom synthesised, oligonucleotide primer. A total of 7 such primers were made, the data from which, in combination with that obtained using universal primer, allowed the derivation of the entire *phoA* sequence. The location of these oligonucleotide primers and their sequences are shown in Figure 5.2. All of the *pho* gene was sequenced on at least one strand and 85% of the sequence on both strands (see Figure 5.2). The complete nucleotide sequence of the *phoA* structural gene, and its 5′- and 3′ non-coding regions is illustrated in Figure 5.3. A revised version of the programme used to predict the secondary structure of staphylococcal SpA was used to predict the secondary structure of the encoded protein (Figure 5.4.)

5.3 DISCUSSION

Nucleotide sequence analysis has shown the *E. coli* JM83 *phoA* gene to be composed of 1,413 bp, encoding for 471 amino acids. The translational initiation codon is preceded by a sequence motif (GGAG) typical of procaryotic ribosome binding (SD) sites. Thus, it is complementary to the nucleotide sequence of the 3′ terminus of the *E. coli* 16S RNA (5′-ACCUCCU-3′) and there is a marked discrimination against the use of C



5.3 Nucleotide Sequence and Deduced Amino Acid Sequence for the Alkaline Phosphatase Gene of *E. coli* JM83

The putative promoter region, SD sequence, signal sequence (s), peptidase cleavage site (c) and transcription termination region are marked. The N-termini of the alkaline phosphatase isoenzymes I_1 and I_2 are shown. The non coding region 5 to the alkaline phosphatase structural gene is numbered according to nucleotide sequence. The *phoA* structural gene is numbered according to its amino acid with the 3 residues of the signal peptide being numbered -1 and the 1st residue of the native protein isoenzyme I being numbered +1.

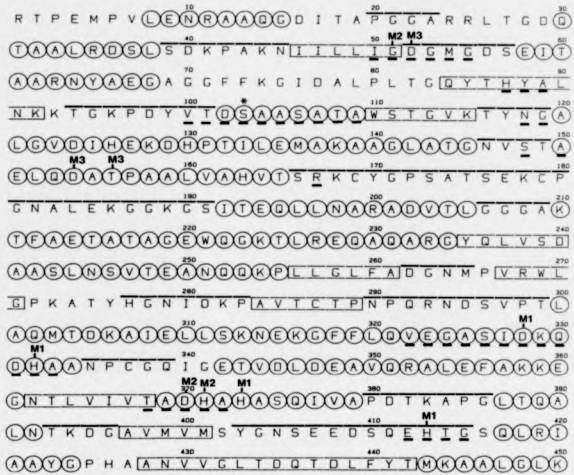


Figure 5.4 Secondary Structure Prediction for *E. coli* Alkaline Phosphatase

The three metal ions contained in the active centre are designated M1, M2 and M3. M2 is the central metal, M1 is above at 3.9 Å and M3 is furthest from the molecular two fold axis. The metal binding ligands are shown with their associated residues. Residues which have the α carbon within 12.5 Å of the central metal and are thought to be involved in the active site are indicated X. The catalytically essential amino acid in the active site a phosphorylatable serine (102) is shown (*), (Wyckoff *et al*, 1983).

nucleotides, and a preference for the use of A, in the immediate flanking regions (Stormo *et al.*, 1982). A hexanucleotide sequence (TATAGT), beginning 52 nucleotides 5′ to the translational initiation codon (position 198 to 203, Figure 5.3) corresponds closely to the recognised consensus sequence of the -10 (Pribnow box) *E. coli* promoter element, TATaaT (where the upper case letters represent the most conserved nucleotides of the promoter site). The distance of this -10 promoter from the start codon is 46 bp, which is similar to that of the *E. coli lacZ* gene (43 bp).

There are two hexanucleotide sequences that closely resemble the consensus *E. coli* ⁻35 promoter sequence (TTGaca). The first, CTGTCA, occurs at nucleotide position 170 to 175, and the second, TTGTCA, at position 181-186 (see Figure 5.3). The first putative ⁻35 promoter sequence differs from the *E. coli* consensus at 2 nucleotide positions, while the second deviates at only one nucleotide position. However, Stefano and Gralla (1982a and b) have shown the nucleotide spacing between the ⁻35 and ⁻10 site of the *lac* promoter to be a crucial factor in promoter strength, with a spacing of 17 nucleotides resulting in maximum promoter strength. Indeed, almost all of the procaryotic promoters examined to date (Rosenberg and Court, 1979; Siebenlist *et al.*, 1980) exhibit a spacing of 17± 1 nucleotide. This would position the ⁻35 site between the two sequences mentioned above at position 175 to 180 (Figure 5.3), the sequence of which, $^{+}$1(C)-ATAAAG-(T)$^{-}$1, exhibits little resemblance to the consensus ⁻35. Other genes have been sequenced that appear not to contain a ⁻35 consensus-like sequence 17 bp 5′ to their ⁻10 motifs (eg., the *cpg* gene; Minton and Clarke, 1985), or have a ⁻35 consensus-like sequence positioned 20 to 23 bp upstream of their ⁻10 sequences (eg., *lac*1, *gal*P1 and; Rosenberg and Court, 1979). Therefore, it is not possible from theoretical study to indicate which region of the *phoA* promoter is acting as the ⁻35 polymerase binding site. It is, however, unlikely to be the second consensus-like sequence, TTGTCA, as this would leave a spacing of only 11 bp between the ⁻35 and ⁻10 promoter regions, and no spacing less than 16 bp has been observed in those promoters examined to date.

The *phoA* gene is immediately followed by a palindromic sequence, consisting of 38 nucleotides, followed by a T-rich stretch of residues, indicative of a transcription terminating region. A possible mRNA hairpin structure for this region is schematically represented in Figure 5.5.

The deduced amino acid sequence of alkaline phosphatase differs from that obtained by protein sequencing (Bradshaw *et al.*, 1981) at only three residues, viz, positions 16 and 36 (Asp for Asn) and 177 (Glu for Gln). These differences are probably due to a misinterpretation of the results of amino acid analysis, as the particular amino acids involved are often difficult to distinguish from one another in the method used. The amino acid sequence deduced here from the nucleotide sequence is more likely to be correct as it is identical to that deduced from the partial nucleotide sequences of other investigators (Boidol *et al.*, 1982; Inouye *et al.*, 1982; Kikuchi *et al.*, 1981), and to that more recently derived from the entire nucleotide sequence of *E. coli* strain 294 (Chang *et al.*, 1986). It would also appear that the sequence of the *phoA* gene is highly conserved between the different *E. coli* strains analysed.

The amino acid sequence of the alkaline phosphatase protein consists of 471 residues. The first 21 amino acids of the N-terminus represent the signal peptide sequence, which is cleaved, following translocation across the cytoplasmic membrane, to leave a mature protein composed of 450 amino acids. The translational initiation codon is slightly unusual, being a GTG (coding for valine) rather than a ATG (coding for methionine). However, it has been shown that the 3 codons capable of initiating the translation of a gene, ATG, GTG and TTG, when positioned at the extreme 5'-end of a structural gene, encode an N-formyl-methionine residue. GTG functions as an initiation codon in 4% of *E. coli* genes, whereas ATG has that role in 91.2% of *E. coli* genes (GenBank, 1984). Between the codon encoding the N-terminal arginine of the native alkaline phosphatase protein and the SD sequence, there are only two codons capable of initiating translation in the correct reading frame. Both of these are GTG, the first at nucleotide position 283

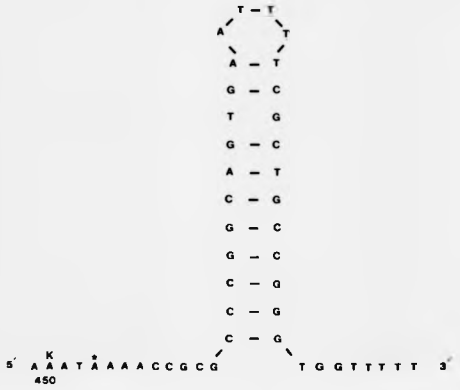


Figure 5.5 **Hypothetical Secondary Structure for the Transcription Termination Sequence in the 3′ Region Flanking the *pho*A Coding Sequence**

Numbered according to the deduced amino acid shown in Figure 5.3. The translation termination codon is denoted by an asterisk (*).

(amino acid -21, relative to the N-terminal Arg) and the second at position 334 (amino acid -4). The former of these is most likely to act as the initiation codon as it is in a suitable position downstream of the SD sequence (9 nucleotides), it is flanked by polyA tracts, a translational termination codon is present in both of the nonsense reading frames, and translation from this codon allows for a polypeptide, 21 amino acids in length, at the N-terminus of the native alkaline phosphatase protein that has the qualities normally attributed to a signal peptide. The properties of this putative signal peptide include: a positively charged N-terminal region (Lys at position $^{-}$20); a hydrophobic core region, 12 amino acids in length (consisting of 75% hydrophobic residues) and possessing a high helix forming potential which ends 4 amino acids prior to the peptidase cleavage site. The amino acids present from $^{-}$5 to $^{+}$5 are consistent with those reported by Perlman and Halvorson (1983) and von Heijne (1983, 1984a) to be common at a peptidase cleavage sites ie., Ala at $^{-}$1 and Pro at $^{-}$5 (more detailed preferences are given in Section 3.3.3). The second possible translational GTG start codon, at position 334, meets none of these requirements.

The most notable features of the amino acid composition *per se* are the high content of Ala (14%) and Gly (10%) residues, the low content of Trp (1%) and Cys (1%) amino acids, and the overall net negative charge of the protein (excess of 12 negative charges per subunit).

Examination of alkaline phosphatase both in solution and in crystal forms by a variety of methods (including X-ray crystallography, X-ray diffraction and NMR; (Wyckoff *et al.*, 1983), combined with knowledge of the amino acid sequence (Bradshaw *et al.*, 1981), has allowed models for a three dimensional structure to be presented (see Figure 5.6). A revised version of the programme used to predict the secondary structure of staphylococcal SpA (Section 3.3.4.2) was used to predict the secondary structure of *E. coli* alkaline phosphatase (Figure 5.4). A comparison of the predicted secondary structure to that deduced by experimentation (Figure 5.6) showed that the computer

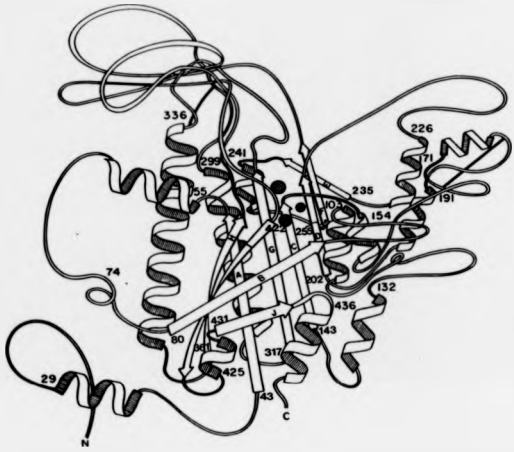


Figure 5.6 Ribbon Diagram of the Complete Monomer of *E. coli* Alkaline Phosphatase

Strands in the sheet are labelled A through J in order along the polypeptide, and their N-terminal amino acid is indicated. Helices are numbered according to their N-terminal amino acid. Adapted from Wyckoff *et al.*, 1983.

accurately predicted 68% of the α-helix and β-sheet regions. Correctly predicted regions included the position of 6 of the 10 β-sheets and 11 of the 14 α-helix regions. Two of the regions not predicted (C and M Figure 5.6) were only 4 residues in length. Of the helical regions predicted using the computer programme, 6 were twice as long as those deduced by experimental methods. However, as Wyckoff *et al.* (1983) stated that the end points of the helical regions are not firmly defined, the computer prediction may be more accurate than it first appears. Such comparisons between a "known" (subject to continuous refinement, specifically in the active site, where substantial conformational flexibility has been observed) structure and that formed using computer predictions, allow the assessment of the value of such programmes for use with proteins for which a nucleotide or amino acid sequence is known, but no experimental data on secondary structure is available.

The active site of a single subunit of alkaline phosphatase consists of three distinct metal binding sites and one phosphorylatable serine (residue 103 which initiates phosphate ester hydrolysis). It has been suggested that the arginine residue at position 167 is directly involved in phosphate hydrolysis (Wyckoff *et al.*, 1983), its positive charge having a stabilising effect on the reaction. However, Butler-Ranshoff *et al.* (1988) have shown that the charge at residue 167 is not required for the hydrolysis of phosphate monoesters, and that its replacement with lysine or glutamine (which are capable of maintaining a hydrogen bond with aspartic acid 102, a bond thought likely to be necessary for the maintenance of serine 103 in a suitable orientation for nucleophilic attack on the phosphate), results in only a slight drop in catalytic activity. Amino acids thought to be involved in the active site, by virtue of close contact (12.5Å) with the central metal group, and the probable bridging ligands are shown in Figure 5.4. These amino acids are in a non-contiguous arrangement with an isoleucine residue being nearest the N-terminus at position 50, and a glycine nearest the C-terminus at position 415. The N-terminal sequence prior to residue 50 has been shown to be important for alkaline phosphatase activity. Cleavage with trypsin removes the first eleven residues of the native protein, resulting in a change in the overall structure of the

molecule concomitant with a 20% reduction in specific activity and an alteration in the affinity for metal ions (Roberts and Chlebowski, 1984, 1985). Cleavage using protease V-8 breaks the bond between the glutamine residue at position 9 and the asparagine residue at position 10. The removal of the first nine amino acids has no dramatic effect on the activity of alkaline phosphatase, whereas the loss of an additional dipeptide (asparagine-10, and arginine-11) results in the aforementioned loss of activity. As a consequence, if the metal ions are removed from the 11 amino acid-truncated form, the resultant monomeric subunits, upon addition of the essential metal ions, are unable to reconstitute to form an active enzyme. Thus, the asparagine-10 and arginine-11 dipeptide appears to be crucial in the assembly of the native enzyme dimer (Tyler-Cross *et al.*, 1989). Further studies by Olafsdottir and Chlebowski (1989), using hybrid dimers where only one of the alkaline phosphatase subunits has been cleaved by trypsin, suggests the presence of a strong sub-unit interaction within the alkaline phosphatase dimer, where an alteration at the arginine-11 and alanine-12 residues can effect the active site on both the truncated and native subunits.

In order to maintain a high degree of catalytic activity in the proposed alkaline phosphatase::SpA and SpA::alkaline phosphatase chimeric proteins, the protein regions forming the active sites, or known to effect the active site, should remain intact. Therefore, the preferred site for the N-terminal fusion of the alkaline phosphatase gene is located in the nucleotide region between nucleotide 346 (first base of triplet coding for arginine 1) and nucleotide 373 (first base of triplet coding for asparagine-10). The preferred site for the C-terminal fusion is in the nucleotide region downstream of the nucleotide triplet coding for glycine-415. The region of the *spa* gene to be used in a fusion must contain at least two of the five IgG binding domains (E, D, A, B, C) in order to maintain a reasonable degree of activity and the N-terminal half must also contain the region coding for the SpA signal peptide.

Examination of the *phoA* and *spa* nucleotide sequences revealed a lack of naturally occurring suitable restriction enzyme cleavage sites in the aforementioned regions. Therefore, it was necessary to introduce suitable restriction enzyme cleavage sites using the technique of oligonucleotide site directed mutagenesis. These sites must fulfil two main criteria. Firstly, they must be compatible with restriction enzyme sites in a multiple site linker that will enable both the *spa* and *phoA* genes to be inserted in the correct order (as the N- or C-terminal half of the fusion) whilst ensuring that the gene forming the C-terminal portion of the fusion is in the same reading frame as the N-terminal gene. Secondly, any restriction enzyme cleavage site introduced should alter the gene sequence to be fused as little as possible. The overall strategy proposed is shown in detail in Figure 5.7. Oligonucleotide mutagenesis and construction of the *phoA*::*spa* gene fusion has been carried out by the Gene Expression Section of the Division of Biotechnology and is currently nearing completion.

Figure 5.7 Oligonucleotide mutations required to allow the Construction of *spa-phoA* Gene Fusions

A. *phoA-spa* fusion

phoA: *Nde*I site created across the translation initiation codon

: *Str*I site created across nucleotides coding for lysine - and the termination codon.

spa: No mutagenic sites needed as naturally occuring *Bcl* I and *Pst*I restriction cleavage sites can be used. The action on *Bcl* I is prevented by methylation of adenine by the *E. coli dam* coded methylase M, therefore the *spa* gene must first be transformed into an *E. coli dam*⁻ strain.

↑ site of cleavage of the restriction enzyme
n nucleotide that has been substituted using Site Directed Mutagenesis
▬ a.a residue change resulting from nucleotide substitution
* mutgenic restriction site to be introduced
⑥ denotes a restriction cleavage site in the *pmt* linker DNA suitable for insertion of *spa* or *phoA* DNA after cleavage with corresponding restriction endonuclease marked ⑥

B. To make *spa-phoA* fusion

spa: *Nde*I site created across the translation initiation codon

*Xho*I site created cross *Pst*I site of *spa* sequence m nt position 2010-2016 (numbering according to Figure 5.2)

PhoA: *Stu*I site created across nucleotides coding for Methionine-5 and proline-6 (numbering according to Figure 5.3)

*Hin*dIII site created across nucleotides 1745-1750, downstream of the *phoA* gene so that non-*phoA* DNA can be eliminated from fusions (numbering according to Figure 5.3).

C.

Construction of fused genes using pIMTL vectors and pIMTL linker (pIMTL20) system (Chambers *et al.*, 1988).

CHAPTER 6

General Discussion

In recent years the applications of SpA in the field of biotechnology have increased dramatically. It was immediately apparent that such a versatile protein would be required in large quantities. Expression of the *spa* gene in *E. coli* had been achieved (Duggleby and Jones, 1983; Uhlen *et al.*, 1983) though the levels of SpA produced did not appear to be higher than those seen in the natural host *S. aureus* Cowan I. As at this time no nucleotide sequence was available and only a partial amino acid sequence of the protein, encompassing the four IgG-binding domains and the repetitive region of the C-terminus had been determined (Sjodahl, 1976; 1977 a & b), it was necessary to determine the entire nucleotide sequence of the *spa* gene. This would allow further studies on the gene and its deduced a.a sequence in order to further enhance the expression of *spa* in *E. coli* and also to allow the construction of precisely defined gene fusions between *spa* and various enzyme genes for use as immunodiagnostic reagents.

The work undertaken in this thesis describes the determination and analysis of the nucleotide sequence and deduced a.a sequence of *spa* from *S. aureus* Cowan I (NCTC8530), the manipulation of the gene to achieve enhanced expression in *E. coli* and the characterisation of the SpA IgG-binding peptides produced.

Analysis of the nucleotide and the deduced amino acid sequence has shown;

i. *spa* encodes a protein, including a signal sequence of 36 a.a, of 508 a.a having a molecular weight of 55,426 D. This protein can be seen to consist of two separate functional regions; an IgG-binding region, and a cell-membrane-and cell-wall-binding region.

ii. A fifth N-terminal region (E), homologous to the four IgG-binding regions reported by Sjodahl (1976, 1977 a & b). This has subsequently been shown to have a weak IgG binding activity (Moks *et al.*, 1986).

iii The C-terminus can be divided into two structurally different regions. The Xr region consists of 10 octapeptide repeats, this region is shown to be hydrophilic and is proposed to act as the cell wall spanning domain of SpA. The Xc region coding for 88 a.a, contains a 20 a.a. hydrophobic region at its C-terminus followed by five charged residues. This region is proposed to bind to the cell membrane and act as an anchor region for the protein A molecule, with the charged tail acting as a stop-transfer signal by binding to the cytoplasmic side of the inner membrane (von Heijne, 1981; Blobel, 1980). This hydrophobic region and a pentapeptide sequence just prior to the hydrophobic sequence have been reported to occur in many Gram-positive proteins which have a membrane binding role (Fischetti, *et al.*, 1990).

iv Studies on internal homologies within the *spa* sequence at both the nucleotide and a.a. level have shown the gene to be composed of a series of repeated sequences, that appear to be more highly conserved at the amino acid level than the nucleotide level, thus showing a strong evolutionary pressure to conserve particular protein sequences. Comparison of this *spa* gene with two other *S. aureus spa* genes suggest that the *spa* gene may have evolved by two independent stepwise gene region duplications. One to form the 5 IgG-binding domain repeat, indicating that 5 IgG-binding domains may represent the maximum number for optimum IgG binding. The other to form the octapeptide repeats of the cell wall binding region, Xr, here the number of octapeptide regions were seen to vary between the *S. aureus* strains (with the lowest number seen for NCTC8530, ten repeats).

v. Analysis of the codon usage of *S. aureus* appeared to indicate a difference in bias depending on whether the gene examined was derived from the chromosome, a plasmid or a transposon; surprisingly *spa* showed a greater similarity to the codon bias seen for plasmid derived genes rather than those from the chromosome. However, most of the other chromosomal genes examined are normally expressed at far lower levels than those seen for *spa* in *S. aureus*. As a difference in codon bias has been shown to exist in *E. coli* between strongly and weakly expressed genes, a similar system could exist in *S. aureus*. Until more of *S. aureus* genes have been analysed no firm

conclusions can be made. A possible explanation (involving the presence in *spa* of a rarely used *E. coli* codon synonym for *leu*) for some of the differences in SpA levels produced between the C-terminally truncated SpA and the full length SpA has been proposed, Chapter 3, Section 3.3.12.

Expression studies of the cloned *spa* gene have revealed:

i. The *spa* promoter and Shine-Dalgarno sequence function efficiently in *E. coli*, achieving up to half the level of SpA seen when expressed from the *E. coli lac* promoter.

ii A four-fold enhancement in SpA levels in *E.coli* was achieved by placing the *E. coli lac* promoter immediately upstream of the *spa* gene.

iii. Scale up of SpA production to 150 and 400 litre batch fermenters has enabled yields of up to 1.5 g of SpA per litre of culture to be achieved. However the recombinant clones are extremely fragile when expressing a high level of SpA. This has caused problems which hopefully should be overcome by using a vector system containing a fully repressible promoter.

iv. The Xc region of the SpA has been shown to affect the cell morphology dramatically when large amounts of SpA are present in the cell. Similar findings have been reported for the over expression of other of heterologous proteins (Schoner *et al.*, 1985; Schoemaker *et al.*, 1985; Georgiou *et al.*, 1986).

v. The lack of the Xc region of the SpA molecule has been shown to affect the membrane binding ability of the truncated protein.

vi. The existence of a possible regulatory control sequence for *spa* expression is proposed. Evidence presented in this thesis suggested this is located in the 608 bp region of *S. aureus* DNA prior to the *spa* structural gene (Chapter 4, Section 4.2.3.4). It was of interest to note the presence of two sequences within this region that possess a strong homology to both the proposed common regulatory site of catabolite repressed *B. subtilis* promoters (Weickert and Chambliss, 1990) and to the CRP binding site and

other known operator sequences of *E. coli* promoters (Studnicka, 1987).

vii. SpA when produced in *E. coli* was shown to be localised to the periplasm of the cell, though traces of the precursor protein form were seen in the membrane fraction of the cell.

viii. Proteolytic degradation of the mature form of SpA was seen to occur when produced in an *E. coli* host. This degradation appeared to be occurring mainly at the C-terminal end of the protein, and appeared to occur at specific sites in the polypeptide rather than a general "chewback" of the end terminus of the protein. Such proteolysis is a common problem in the expression of heterologous proteins in bacteria (Uhlen and Moks, 1990). It has also recently been reported that the susceptibility of a recombinant protein to proteolysis can be affected by its neighbouring domains (Murby *et al.*, 1991).

The results presented in this thesis provide a substantial basis for further work in several areas such as control of expression of *spa* either using an alternative *E. coli* vector system or perhaps eventually using its own regulatory controls. Stringent control of SpA production may obviate the problems of host cell fragility seen in this study and so increase the amount of SpA recovered from large-scale fermentations. The nucleotide sequence of the gene *spa* Cowan I (NCTC8350) and of the other *spa* genes published during the course of this thesis (Uhlen, *et al.*, 1984a; Colbert and Anilonis (1984) have been used the construction of gene fusions between *spa* and the genes of many enzymes (see Appendix III).

REFERENCES

Abrahmsen, L., Moks, T., Nilsson, B. and Uhlen, M. (1986). Nucleic Acids Res., **14**, 6487-7500.

Anilionis, A. and Palmer, J.L. Beta-glucuronidase-S.aureus protein A fusion Patent. World Patent (WPI) no. 87-251144136. Patent Assignie : Repligen Corp.

Atkins, J.T. (1979). Nucleic Acids Res., **7**, 1035-1041.

Baldwin, R.W., Price, M.R. and Robins, R.A. (1973). Br. J. Cancer (Suppl.) **28**, 37-47.

Barkas, T. (1981). J. Clin. Lab. Immunol, **5**, 59

Baron, S.F., Williams, D.S., May, H.D., Patal, P.S., Aldrach, H.C. and Ferry, J.G. (1989). Arch. Microbiol., **151**, 307-313.

Bayer, M.H., Costello, G.P. and Bayer, M.E. (1982). J. Bacteriol., **149**, 758-767.

Bendayan, M. (1984). **1**, p349-372.

Berg, P.E. (1981). **146**, 660-667.

Berzofsky, J.A. (1985). Science, **229**, p932-940.

Biodol, W., Simonis, M., Topert, M. and Siewert, G. (1982). Mol. Gen. Genet., **185**, 510-512.

Birnboim, H.C.and Doly, J. (1979). Nucleic Acids Res., **7**, 1513-1523.

Bjorck, L., Peterson, B., and Sjoquist, J. (1972). Eur. J. Biochem., **29**, 579-584.

Bjorck, L., Kastern, W., Lindahls, G., and Wideback, K. (1987). Molecular Immunology, **24**, 1113-1122.

Blobel, G. (1980). Proc. Natl. Acad. Sci., USA., **77**, 1496-1500.

Blumenthal, H. J. (1972). 111-135. In J.O.Cohen (ed). The Staphylococci. John Wiley & Sons, Inc. New York.

Bradshaw, R.A., Cancedda, F., Ericsson, L.H., Neumann, P.D., Piccoli, S.P., Schlesinger, M.J., Shriefer, K. and Walsh, K.A. (1981). Proc. Natl. Acad. Sci, USA., **78**, 3473-3477.

Brisson-Noel, A. and Courvalin, P. (1986) Gene, **43**, 247-253.

Brosius, J. (1984). Gene, **27**, 161-172.

Brown, J.P., Tamerius, J.D. and Hellstrom, I. (1979). J. Immunol. Methods, **31**, 201-209.

Butler-Ranshoff, J.E., Kendall, D.A. and Kaiser, E.T. (1988). Proc. Natl. Acad. Sci., USA., **85**, 4276-4278.

Carter, P., Bedouelle, H. and Winter, G. (1985). Nuc. Acids. Res., **13**, 4431-4443.

Chambers, S.P., Prior, S.P., Barstow, D.A. and Minton, N.P. (1988). Gene **68**, 139-149.

Chang, C.N., Inouye, H., Model, P. and Beckwith, J. (1980). J. Bacteriol., **142**, 251-255.

Chang, C.N., Kuang, W. and Chen, E.Y. (1986). Gene, **44**, 121-125.

Cheng, Y-S.E., Kwoh, D.Y., Kwoh, T.J., Saltvedt, B.C. and Zipser, D. (1981). Gene, **14**, 121-130.

Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzym. **47**, 45-148.

Clark, B.F.C., and Marcker, K.A. (1966). J. Mol. Biol., **17**, 394-406.

Clewell,D.B. and Helsinki, D.R. (1969). Proc. Natl. Acad. Sci., USA., **62**, 1159-1166.

Cohen, D., Fer, M.F., Pearson, J.W. and Herberman, R.B. (1984). J. Biol. Resp. Modif., **3**, 271-277.

Colbert, D. A. and Anilionis, A.: European Patent Application. (1984). No. A2 0 107509.

Colbert, D., Anilionis, A., Gelep, P., Farley, J and Breyer, R. (1984). J. Biol. Resp. Modif., **3**, 255-259.

Coley, W.B. (1906). Am. J. Med. Sci., **131** 375.

Cornelis, P., Digneffe, C. and Willemot, K. (1982). Mol. Gen. Genet., **186**, 507-511.

Davis, B.D., Dulbecco, R., Eisen, H.N. and Ginsberg, H.S. (1980). Microbiology 3rd Edition (Harper and Row, USA).

Decrombrugghe, B. and Pastan, I. (1978). The Operon. Ed. Miller, J.H. and Reznikoff, W.S. Cold Spring Harbor Labo. 303-324 and 449.

Deisenhofer, J. (1981). Biochemistry, **20** 2361-2370.

Deisenhofer, J., Jones, T.A., Huber, R., Sjodahl, J. and Sjoquist, J. (1978). Hoppe Seyler's Z. Physiol. Chem., **359**, 975.

Deniniger, P.L. (1983). Anal. Biochem., **129**, 216-223.

Duggleby, C.J. and Jones, S.A. (1983). Nucleic Acids Res., **11**, 3065-3075.

Ebright, R.H. (1982). Molecular Structure and Biological Activity, Elsevier, New York, USA 91-100.

Edelman, G. M., Cunningham, B.A., Gall, W.E., Gottlieb, P.D., Rutishauser, U. and Waxdal, M.J. (1969). Biochem. **63**, 78-85.

Edmann, P., and Begg, G. (1967). Europ. J. Biochem. **1**, 80-91.

Ehrlich, S.D., Niaudet, B. and Michel, B. (1982). Proc. Natl. Acad. Sci, USA., **96**, 19-29.

Emr, S.D., Schwartz, M. and Silhavy, T.J. (1978). Curr. Top. Microbiol. Immunol., **75**, 5802-5806.

Emr, S.D., Hall, M.N. and Silhavy, T.S. (1980). J. Cell Biol., **86**, 701-711.

Endresen, C. (1979). Acta. Path. Microbiol. Scand. Sect. C., **87**, 185-189.

Endresen, C. and Grov, A. (1976). Acta. Path. Microbiol. Scand. Sect. C, **84**, 397-402.

Fahnestock, S.R. (1987). Trends Biotechnol. **5**, 79-83.

Fahnestock, S.R., Saunders, G.W., Guyer, M.S., Lofdahl, S., Guss, B., Uhlen, M. and Lindberg, M. (1986a). J. Bact., **165**, 1011-1014.

Fahnestock, S.R., Alexander, P., Nagle, J. and Filpula, D. (1986b). J. Bact., **167**, 870-880.

Fairweather, N., Kennedy, S., Foster, T.J., Kehoe, M. and Dougan, G. (1983). Infect. Immun., **41**, 1112-1117.

Felmlee, T., Pellett, S., Lee, E-Y. and Welch, R.A. (1985a). J. Bacteriol., **163**, 88-93.

Felmlee, T., Pellett, S., Welch, R. (1985b). J. Bacteriol. **163**, 94-105.

Fer, M.F. and Oldham, R.K. (1985). Contemporary Topics in Immunology. Volume 15: Immune Complexes and Human Cancer (Salinas, F.A. and Hanna, M.G., Jr., Eds.) Plenum Press, New York., **15**, 257-276.

Fer, M.F., Berman, J.A., Stevenson, H.C., Maluish, A., Moratz, C., Delawter, T., Foon, K., Herberman, R.B., Oldham, R.K., Terman, D.S., Young, J.B. and Daskai, Y. (1984). J. Biol. Resp. Modif. **3**, 352-358.

Fischetti, V.A., Pancholi, V. and Schneewind, O. (1990). Mol. Micro. **4**, 1603-1605.

Forsgren, A. (1969). Acta. Pathol. Microbiol. Scand., **75**, 481-490.

Forsgren, A. (1970). Infect. Immun., **2**, 672-673.

Forsgren, A. (1972). Acta. Pathol. Microbiol. Scand. Sect. B, **80**, 564-570.

Forsgren, A. and Sjoquist, J. (1966). J. Immunology, **97**, 822-827.

Forsgren, A. and Sjoquist, J. (1969). Acta Pathol. Microbiol Scand., **75**, 466-480.

Foresgren, A., Ghetie, V., Lindmark, R. and Sjoquist, J. (1983). Staphylococci and Staphylococcal Infections. Academic Press Inc. (London) Ltd. Easmon, C.S.E. and Adlam, C., **2**, 429-480.

Friskensky, B., Lis, J. and Wu, R. (1982). Nucleic Acids Res., **10**, 6451-6463.

Gait, M.J., Matthes, H.W.D., Singh, M., Sproat, B.S. and Titmus, R.C. (1982). Nucleic Acids Res., **10**, 6243-6254.

Garen, A. and Levinthal, C. (1960). Biochim. Biophys. Acta., **38**, 470-483.

Genbank (1984). Nucloetide Sequences. Nucleic Acids Res. Supplement parts 1 and 2.

Georgiou, G., Telford, J.N., Shuler, M.L. and Wilson, D.B. (1986). Appl. Environ. Microbiol., **52**, 1157-1161.

Georgiou, G. and Baneyx, F. (1990). Ann. N.Y. Acad. Sci., **589**, 139-147.

Ghetie, V., Nilsson, K. and Sjoquist, J. (1974). Eur. J. Immunol., **4**, 500-505.

Gilbert, H.J., Blazek, R., Bullman, H.M.S., Minton, N.P. (1986). J. Gen. Micro., **132**, 151-160.

Gillam, S., and Smith, M. (1979). Gene, **1**, 81-97.

Goding, J.W. (1978). J. Immun. Meths., **20**, 241-253.

Gold, L., Pribnow, D., Scheider, T., Shinedling, S., Singer, B.S. and Stormo, G. (1981). Ann. Rev. Microbiol., **35**, 265-403.

Gordon, B.R., Matus, R.E., Saal, S.D., MacEwen, E.G., Hurwitz, I., Stenzel, K.H. and Rubin, A.L. (1983). J. Natl. Cancer Inst., **70**, 1127-1133.

Gotz, F., Ahrne, S. and Lindberg, M. (1981). J. Bacteriology, **145**, 74-81.

Gotz, F., Popp, F., Korn, E. and Schleifer, K.H. (1985). Nucleic Acids Res., **13**, 5895-5906.

Gouy, M. and Gautier. (1982). Nucleic Acids Res., **10**, 7055-7073.

Gray, G.S. and Kehoe, M. (1986). Immunol. Rev., **46**, 615-618.

Greenwood, F.C., Hunter, W.M. and Glover, J.S. (1963). Biochem. J., **89**, 114-119.

Gronowiz, E., Coutinho, A. and Melchers, F. (1976). Eur. J. Immunol., **6**, 588-590.

Grosjean, H. and Fiers, W. (1982). Gene, **18**, 199-209.

Grov, A. (1967). Acta. Path. Microbiol. Scand., **69**, 567-575.

Grov, A., Myklestad, B. and Oeding, P. (1964). Acta Pathol. Microbiol Scand., **61**, 588-596.

Grunstein, M. and Hogness, D. (1975). Proc. Natl. Acad. Sci,. USA, **72**, 3961-3965.

Gruss, A. and Ehrlich, S.D. (1989). Microbiol. Rev., **53**, 231-241.

Gryczan, T.J. (1982). The Molecular Biology of the Bacilli. Ed. D.A. Dubnau, Academic Press, Inc., New York. Vol. 1, 307-329.

Gryczan, T.J. and Dubnau, D. (1978). Proc. Natl. Acad. Sci., USA, **75**, 1428-1432.

Guss, B., Uhlen, M., Nilsson, B., Lindberg, M., Sjoquist, J. and Sjodahl, J. (1984a). Eur. J. Biochem., **138**, 413-420.

Guss, B., Uhlen, M., Nilsson, B., Lindberg, M., Sjoquist and Sjodahl, J. (1984b). Eur. J. Biochem., **143**, 685.

Guss, B., Leander, K., Hellman, U., Uhlen, M., Sjöquist, J., Lindberg, M. (1985). Eur. J. Biochem., **153**, 579-585.

Guss, B., Eliasson, M., Olsson, A., Uhlen, M., Frj, A., Jornvall, H., Janingmar, F. and Lindberg, M. (1986). EMBO, **5** 1567-1575.

Hanson, D.C. and Schumaker, V.N. (1984). J. Immunol., **132**, 1397-1409.

Hellebust, H., Murby, M., Abrahmsen, L., Uhlen, M. and Enfors, S-O., (1987). Bio/Technol., **7**, 165-168.

Hellebust, H., Veide, A. and Enfors, S-O. (1988). J. Biotech., **7**, 185-198.

Hellstrom, I. and Hellstrom, K.E. (1970). Int. J. Cancer, **5**, 195-201.

Hellstrom, I., Hellstrom, KE., Evans, C.A., Heppner, G.H., Pierce, G.E. and Yang, J.P.S. (1969). Proc. Natl. Acad. Sci., USA, **62**, 362-368.

Henner, D.J. and Hoch, J.A. (1980). Microbiol. Rev., **44**, 57-82.

Heppel, L.A. (1967). Science, **156**, 1451-1455.

Hibbs, M.L., Walker, I.D., Kirszbaum, L., Pietersz, G.A., Deacon, N.J., Chambers, G.W., McKenzie, I.F.C. and Hogarth, P.M. (1986). Proc. Natl. Acad. Sci., **83**, 6980-6984.

Hirst, T.R., Randall, L.L. and Hardy, S.J.S. (1984). J. Bacteriol., **157**, 637-642.

Hjelm, H., Hjelm, K. and Sjoquist, J. (1972). FEBS Lett., **28**, 73-76.

Hjelm, H., Sjodahl. J. and Sjoquist, J. (1975). Eur. J. Biochem., **57**, 395-403.

Hollingshead, S.K., Fischetti, V.A. and Scott, J.R. (1986). J. Biol. Chem. **261**, 1677-1686.

Holmes, D.S. and Quigley, M. (1981). Anal. Biochem., **114**, 193-197.

Hopp, T.P. (1986). J. Immunol. Methods, **88**, 1-18.

Hopp, T.P. and Woods, K.R. (1981). Proc. Natl. Acad. Sci. U.S.A., **78**, 3824-3828.

Hunkapillar, M.W., Hewick, R.M., Dreyer, W.J. and Hood, L.E. (1983). Meths. Enzymol., **91**, 399-413.

Ikemura, T. (1981b). J. Mol. Biol., **151**. 389-409.

Ikemura, T. (1981a). J. Mol. Biol., **146**, 1-21

Ikemuar, T. and Ozeki, H. (1982). **47**, 1087-1087.

Ikemura, T (1985). Mol. Biol. Evol., **2**, 13-34.

Inganäs, M., Johansson, S.G.O. and Bennich, H.H. (1980). Scand. J. Immunol., **12**, 23-31.

Inganäs, M., Johansson, S.G.O., Sjoquist, J. (1981). Scand. J. Imunol., **14**, 379-388.

Inganäs, M and Nilsson, U.R. (1981). ACTA Universitatis Upsaliensis, Upssala, Faculty of Science.

Innis, M.A., Tokunaga, M., Williams, M.E., Loranger, J.M., Chang, S-Y., Chang, S. and Wu, H.C. (1984). Proc. Natl. Acad. Sci., USA, **81**, 3708-3712.

Inouye, H., Barnes, W. and Beckwith, J. (1982). J. Bact., **149**, 434-439.

Inouye, H., Michaelis, S., Wright, A. and Beckwith, J. (1981). J. Bacteriol., **146**, 668-675.

Jackson, M.P. and Iandolo, J.J. (1986). J. Bact., **167**, 726-729.

Janniére, L., Bruand, C. and Ehrlich, S.D. (1990). Gene, **87**, 53-61.

Janzon, L. and Arvidson, S. (1990). EMBO J., **9**, 1391-1399.

Janzon, L., Lofdahl, S. and Arvidson, S. (1986). FEMS Microbiol. Lett. **33**, 193-198.

Janzon, L., Lofdahl, S. and Arvidson, S. (1989). Mol. Gen. Genet., **219**, 480-485.

Jensen, K (1958a). Acta Pathol. Microbiol. Scand., **44**, 421-428.

Jensen, K, (1958b). Acta Allergol., **12**, 210-211.

Jensen, K. (1959). Acta Allergol, **13**, 89-100.

Jones, C.L. and Khan, S.A. (1986). J. Bact., **166**, 29-33.

Jones, F.R., Balint, Jr, J.P. and Synder, Jr., H.W., (1986). Plasma Ther. Transfus. Technol., **7**, 333-349.

Jonsson, P., Lindberg, M., Haraldsson, I. and Wadstrom, T. (1985). Infect. and Immun., **49**, 765-769.

Kelley, P.M., Neumann, P.A., Shriefer, K., Cancedda, T., Schlesinger, M.J. and Bradshaw, R.A. (1973). Biochemistry, **12**, 3499-3503.

Kessler, S.W. (1975). J. Immunol., **115**, 1617-1624.

Kikuchi, Y., Yoda, K., Yamasaki, M. and Tamura, G. (1981). Nucleic Acids Res., **9**, 5671-5679.

Kim, J-H. and Weaver, R.F. (1988). Gene, **68**. 315-321.

Kleid, D.G., Yansura, D., Small, B., Dowbenko, D., Moore, D.M., Grubman, M.J., Mckercher, P.D., Morgan, D.O., Robertson, B.H. and Bachrach, H.L. (1981). Science, **214**, 1125-1129.

Klier, A., Fargette, F., Ribier, J. and Rapoport, G. (1982). EMBO J., **1**, 791-799.

Kobatake, E., Ikariyama, Y., Aizawa, M., Miwa, K. and Kato, S. (1990). J. Biotech., **16**, 87-96.

Koch. (1878). Unattributed reference: Davis *et al.*, 1980.

Konigsberg, W. and Godson, L.N. (1983). Proc. Natl. Acad. Sci., USA, **80**, 687-691.

Kreft, J., Parrisius, J., Burger, K.J. and Goebel, W. (1982). IN: "Molecular cloning and Gene Regulation in Bacilli" (Eds. Ganesan, A.T., Chang, S. and Hoch, J.A.) Academic Press, New York, 145-157.

Kronvall, G. (1973). Scand. J. Immunol., **2**, 31-36.

Kronvall, G., and Frommel, D. (1970). Immunochemistry, **7**, 124-127.

Kronvall, G., Seal, U.S., Finstad, J. and Williams, Jr, R.C. (1970). J. Immunol., **104**, 140-147.

Kronvall, G., Seal, U.S., Svensson, S. and Williams, R.C. Jr. (1974). Acta Path. Microbiol. Scand. Sect. B, **82**, 12-18.

Kushner, S.R. (1978). Genetic Engineering (Boyer, H.W. and Nicosia, S. Eds). Elsevier/North Holland Biomedical Press, 17-23.

Kyte, J. and Doolittle, R.F. (1982). J. Mol. Biol., **157**, 105-132.

Laemmli, U.K. (1970). Nature, **227**, 680-685.

Lachia, R.V.F., Genigeorgis, C.A. and Hoeprich, P.D. (1979). J. Clin. Microbiol., **10**, 752-753.

Langone, J.J. (1982a). Adv. in Immunol., **32**, 157-252.

Langone, J.J. (1982b). J. Immunol. Meths., **51**, 3-22.

Langone, J.J., Boyle, M.D.P. and Borsos, T. (1978a). J. Immunol., **121**, 327-332.

Langone, J.J., Boyle, M.D.P. and Borsos, T. (1978b). J. Immunol., **121**, 333-338.

Larson, R. and Messing, J (1983). DNA., **2**, 31-35.

Laskey, R.A., and Mills, A.D. (1977). FEBS Lett., **22**, 314-316.

Lathe, R., Kieny, M.P., Skory, S. and Lecocq, J.P. (1984). DNA, **3**, 173-182.

Laurell, C.B. (1965). Anal.Biochem., **10**, 358-361.

Lew, A.M., Beck, D.J. and Thomas, L.M. (1991). J. Immunol. Meths., **136**, 211-219.

Lewis, V.A., Koch, T., Plutner, H. and Mellman, I. (1986). Nature, **324**, 372-375.

Lind, L (1972). Acta Pathol. Microbiol. Scand. Sect. B, **80**, 702-708.

Lindbladh, C., Mosbach, K. and Bulow, L. (1991). J. Immunol. Meths. **137**, 199-207.

Lindmark, R., Movitz, J. and Sjoquist, J. (1977). Eur. J. Biochem., **74**, 623-628.

Lofdahl, S., Guss, B., Uhlen, M., Philipson, L. and Lindberg, M. (1983). Proc. Natl. Acad. Sci., USA, **80**, 697-701.

Lofkvist, T. and Sjoquist, J. (1962). Acta Pathol. Microbiol. Scand., **56**, 295-304.

Lofkvist, T. and Sjoquist, J. (1963). Int. Arch. Allergy, **23**, 289.

Lowenadler, B., Nilsson, B., Abrahmsen, L., Moks, T., Ljungquist, L., Holmgren, E., Paleus, S., Josephson, S., Philipson, L. and Uhlen, M. (1986). EMBO J., **5**, 2393-2398.

Lowenadler, B., Sansson, B., Paleus, S., Holmgren, E., Nilsson, B., Moks, T., Palm, G., Josephson, S., Philipson, L. and Uhlen, M. (1987). Gene, **58**, 87-97.

Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951). J. Biol. Chem., **193**, 265-275.

Mackintosh, F.R., Bennett, K. and Hall, S.W. (1985). Contemporary Topics in Immunology. Vol. 15. Immune Complexes and Human Cancer (Salinas, F.A. and Hann, M.G. Jr. Eds). Plenum Press, New York, **15**, 239-256.

Mackman, N., Nicaud, J.M., Gray, L. and Holland, I.B. (1985). Mol. Gen. Genet., **201**, 529-536.

Malan, T.P. and McClure, W.R. (1984). Cell, **39**, 173-180.

Maniatis, T., Fritsch, G.F. and Sambrook, J. (1982). Molecular Cloning : A laboratory manual. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

Marmur, J. (1961). J. Mol. Biol., **3**, 208-218.

Maxam, A.M., and Gilbert, W. (1977). Proc Natl. Acad. Sci. U.S.A., **74**, 560-564.

McClatchy, J.K. and Rosenblum, E.D. (1963). J.Bacteriol., **86**, 1211-1215.

McDonnell, M.W., Simon, M.N. and Studier, F.W. (1977). J. Mol. Biol., **110**, 119-146.

Melling, J. and Scott, G.K. (1972). Biochem. J. **130**, 55-62.

Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977). Proc. Natl. Acad. Sci., USA, **74**, 3642-3646.

Messing, J. and Vieira, J. (1982). Gene, **19**, 269-276.

Michaelis, S. and Beckwith, J. (1982). Ann. Rev. Microbiol., **36**, 435-465.

Miller, W.J., Branda, R.F., Hurf, D.D., Wachsman, W., Nelson, N.L. and Jacob, H.S. (1982). Blood, **59**, 1344-1347.

Minton, N.P. and Clarke, L.E. (1985). J. Mol. Appl. Genet., **3**, 26-35.

Moks, T., Abrahmsen, L., Nilsson, B., Hellman, U., Sjoquist, J. and Uhlen, M. (1986). Eur. J. Biochem., **156**, 637-643.

Moore, K.G. and Miura, S. (1987). J. Biol. Chem., **262**, 8806-8813.

Morgan, M.R.A., Johnson, P.M. and Dean, P.D.G. (1978). J. Immunol. Methods, **23**, 381-387.

Mota, G, Ghetie, V and Sjoquist, J. (1978). Immunochemistry, **15**, 639-642.

Movitz, J. (1976). Eur. J. Biochem., **68**, 291-299.

Movitz, J. (1974). Eur. J. Biochem., **48**, 131-136.

Muller, H.P., Schaeg, W. and Blobel, H. (1981). Zbl. Bakt. Hyg. **249**, 443-451.

Muller-Hill, B., Beyreuther, K. and Gilbert, W. (1971). Methods in Enzymology, **XXI**, Part D: Nucleic Acids, **39**, 483-487.

Muller-Hill, B., Crapo, L. and Gilbert, W. (1968). Proc. Natl. Acad. Sci., USA, **59**, 1259-1264.

Munson, L.M., Stormo, G.D., Niece, R.L. and Reznikoff, W.S. (1984). J. Mol. Biol., **177**, 663-684.

Murby, M., and Nygren, P-A., Rondahl, H., Hellman, U., Enfors, S-O. and Uhlen, M. (1991). Eur. J. Biochem. **199**, 41-46.

Murray, C.L., and Rabinowitz, J.C. (1982). J. Biol. Chem., **257**, 1053-1062.

Murphy, E., Huwyler, L. and Do Carmo De Freire. and Bastos, M. (1985). EMBO J., **4**, 3357-3365.

Nichols, B.P., Miozzari, G.F., van Cleemput, M., Bennett, G.N. and Yanofsky, C. (1980). J. Mol. Biol., **142**, 503-517.

Nilsson, B. and Abrahmsen, L. (1990). Methods in Enzymology, **185**, 144-161.

Nilsson, B., Abrahmsen, L. and Uhlen, M. (1985a). EMBO J., **4**, 1075-1080.

Nilsson, B., Holmgren, E., Josephson, S., Gatenbeck, S., Philipson, L. and Uhlen, M. (1985b). Nucleic Acids Res. **13**, 1151-1162.

Nilsson, I., Jonssen, S., Sundqvist, S-B., Ahlberg, A. and Bergentz, S-E. (1981). Blood, **58**, 38-44.

Novick, R., Kornblum, J., Kreiswith, B., Projan, S. and Ross,H. (1989). In Butler, L.O., Harwood, C. and Mosely, B.E.B. (eds). Genetic transformation and expression. Intercept Ltd. Winborn, Dorset, U.K. p495-510.

Ogasawara, N. (1985). Gene, **40**, 145-150.

Olafsdottir, S. and Chlebowski, J.F. (1989). J. Biol. Chem., **264**, 4529-4535.

Oliver, D. (1985). Ann. Rev. Microbiol., **39**, 615-648.

Oskouian, B. and Stewart, G.C. (1990). J. Bact., **172**, 3804-3812.

Osland, A. (1981). Acta. Pathol. Microbiol. Immunol. Scand. Sect. B, **90**, 197-203.

O'Farrell, P (1981). Bethesda Research Labs., Focus, **3**, 1-4.

Pancholi, V. and Fischetti, V.A. (1988). J. Bact., **170**, 2618-2624.

Peng, H-L., Novick, R.P., Kreiswirth, B., Kornblum, J. and Schlievert, P. (1988). J. Bact. **170**, 4365-4372.

Perlman, D. and Halvorson, H.O. (1983). J. Mol. Biol., **167**, 391-409.

Pierschbacher, M., Hayman, E.G. and Ruoslahti, E. (1983). Proc. Natl. Acad. Sci., USA, **80**, 1224-1227.

Pierschbacher, M.D. and Ruoslahti, E. (1984a). Nature, **309**, 30-33.

Pierschbacher, M.D. and Ruoslahti, E. (1984b). Proc. Natl. Acad. Sci., USA, **81**, 5985-5988.

Pribnow, D. (1975). Proc. Natl. Acad. Sci., USA, **72**, 784-788.

Priest, F.G. (1977). Bact. Rev., **41**, 711-753.

Pugsley, A.P. and Schwartz, M. (1984). EMBO J., **3**, 2393-2397.

Raibaud, O. and Schwartz, M. (1984). Ann. Rev. Genet., **18**, 173-206.

Ray, P.K. and Bandyopadhyay, S. (1983). Immunol. Commun., **12**, 453-464.

Recsei, P.A., Gruss, A.D. and Novick, R.P. (1987). Proc. Natl. Acad. Sci., USA, **84**, 1127-1131.

Recsei, P., Kreiswirth, B., O'Reilly, M., Schlievert, P., Gruss, A. and Novick, R.P. (1986). Mol. Gen. Genet., **202**, 58-61.

Reddy, P., Peterkufsky, A. and McKenney, K. (1985). Proc. Natl. Acad. Sci., USA, **82**, 5656-5660.

Reid, T.W. and Wilson, I.B. (1971). Enzymes, **4**, 373-415.

Reynolds, E.S. (1963). J. Cell Biol., **17**, 208-212.

Reznikoff, W.S., Siegele, D.A. Cowing, D. and Gross, C.A. (1985). Ann. Rev. Genet., **19**, 355-387.

Roberts, C.H. and Chlebowski, J.F. (1984). J. Biol. Chem., **259**, 729-733.

Roberts, C.H. and Chlebowski, J.F. (1985). J. Biol. Chem., **260**, 7557-7561.

Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, C. (1984). Nucleic Acids Res., **12**, 6663-6671.

Rodriguez, R.L. and Tait, R.C. (1983). Recombinant DNA Techniques : An Introduction. Publ. Addison-Wesley, Cambridge, Massachusetts, USA.

Romagnani, S., Del Prete, G.F., Maggi, E., Falagiani, P. and Ricci, M. (1980). Immunology, **41**, 921-927.

Rose, J.K. and Shafferman, A. (1981). Proc. Natl. Acad. Sci., USA, **78,** 6670-6674.

Rusenbach. (1884). Unattributed reference: Davis *et al.* 1980.

Rosenburg, M. and Court. D. (1979). Ann. Rev. Genet., **13,** 319-353.

Ruckdeschel, J.C., Codish, S.D., Stranahan, A. and McKneally, M.F. (1972). N. Eng. J. Med., **287,** 1013-1017.

Sako, T., and Tsuchida, N. (1983). Nucleic Acids Res., **11,** 7679-7693.

Sanger, F. and Coulson, A.R. (1978). Febs Lett. **87,** 107-110.

Sanger, F., Nicklen, A.R., and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. U.S.A., **74,** 5463-5467.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H and Rog, R.A. (1980). J. Mol Biol., **143,** 161-178.

Sarvas, M.O. and Palva, I.A. (1983). J. Bact., **155,** 657-663.

Saunders, C.W., Schmidt, G.J., Mirot, M.S., Thomson, L.D. and Guyer, M.S (1984). J. Bacteriol., **157,** 718-726.

Schaller, H., Gray, C. and Herrmann, K. (1975). Proc. Natl. Acad. Sci., USA, **72,** 737-741.

Schatz, P.J. and Beckwith, J. (1990). Annu. Rev. Genet., **24,** 215-248.

Schleifer, K.H. and Kloos, W.E. (1975). Int. J. Syst. Bacteriol., **25,** 50-61.

Schlesinger, M.J., Block, W. and Kelley, P.M (1975). Isozyme I: Molecular Structure. Ed Markert, C., Academic Press, New York, USA, 333-342.

Schoner, R.G., Ellis, L.F., Schoner, B.E. (1985). Bio/Technology, **4**, 151-154.

Schoemaker, J.M., Brasnett, A.H. and Marston, F.A.E. (1985). EMBO J., **4**, 775-780.

Sheagren, J.N. (1984a). N. Eng. J. Med., **310**, 1368-1373.

Sheagren, J.N. (1984b). N. Eng. J. Med., **310**, 1437-1442.

Shepherd, J.C.W. (1984). Trends Biochem. Sci. **9**, 8-10.

Shields, D.C. and Sharp, P.M. (1987). Nucleic Acids Res., **15**, 8023-8040.

Shibata, T., Dasgupta, C., Cunningham, R.P. and Radding, C.M. (1979). Proc. Natl. Acad. Sci., USA, **76**, 1638-1642.

Shine, J. and Dalgarno, L. (1975). Nature, **254**, 34-38.

Siebenlist, U., Simpson, R.B and Gilbert, W. (1980). Cell, **20**, 269-281.

Sjodahl, J. (1976). FEBS Letts., **67**, 62-67.

Sjodahl, J (1977a). Eur. J. Biochem., **73**, 343-351.

Sjodahl, J (1977b). Eur. J. Biochem., **78**, 471-490.

Sjoholm, I. (1975). Eur. J. Biochem., **51**, 55-61.

Sjoquist, J., Meloun, B. and Hjelm, H. (1972a). Eur. J. Biochem., **29**, 572-578.

Sjoquist, J., Movitz, J., Johansson, I. and Hjelm, H. (1972b). Eur. J. Biochem. **30**, 190-194.

Sjorgren, H.O., Hellstrom, I., Bansal, S.C. and Hellstrom, K.E. (1971). Proc. Natl. Acad. Sci., USA, **68**, 1372-1375.

Smith, M. (1986). Ann. Rev. Genet., **19**, 423-462.

Soberon, X., Covarrubias, L. and Bolivar, F. (1980). Gene, **9**, 287-305.

Solal-Celigny, P., Simeon, J., Herrera, A. and Boivin, P. (1985). Biomedicine and Pharmacotherapy, **39**, 177-186.

Southern, E. (1975). J. Mol. Biol., **98**, 503-507.

Staden, R. (1980). Nucleic Acids Res., **8**, 3673-3694.

Steele, G. Jr., Ankerst, J. and Sjogren, H.O. (1974). Int. J. Cancer, **14**, 83-92.

Stefano, J.E. and Gralla, J.D. (1982a). Promoters, structures and function. Eds. Rodriguez, R.L. and Chamberlin, M.J. (Publ. Praeger), 69-79.

Stefano, J.E. and Gralla, J.D. (1982b). Proc. Natl. Acad. Sci., USA, **79**, 1069-1072.

Stevens, W.J. and Bridts, C. (1981). Immunol. Lett., **3**, 1-

Stewart, G.A., Varro, R. and Stanworth, D.R. (1978). Immunology, **35**, 785-791.

Stormo, G.D., Schneider, T.D. and Gold, L.M. (1982). Nucleic Acids Res., **10**, 2971-2996.

Studnicka, G.M. (1987). Gene, **58**, 45-57.

Sukumar, S., Zbar, B., Terata, N. and Langone, J. (1984). J. Biol. Resp. Modif., **3**, 303-315.

Surolia, A., Pain, D. and Khan, I. (1982). TIBS, **7**, 74-76.

Suzuki, S., Oldberg, A., Hayman, E.G., Pierschbacher, M.D. and Ruoslahti, E. (1985). EMBO J., **4**, 2519-2524.

Swamy, K.H.S. and Goldberg, A.L. (1981). Nature, **292**, 652-654.

Swanstrom, B. and Shank, P.R. (1978). Anal. Biochem., **86**, 184-192.

Szostak, J.W., Stiles, J.I., Tye, B-K, Chill, P., Sherman, F. and Wu, R. (1979). Methods in Enzymology, **68**, 419-428.

Takita, H. (1970). J. Thorac. Cariovasc. Surg., **59**, 642.

Talmadge, K.E. and Gilbert, W. (1982). Proc. Natl. Acad. Sci., USA, **79**, 1830-1833.

Tarkhanova, I.A., Rudenskii, A. Yu., Sycheva, I.M. and Kul'berg, A. Ya. (1980). Plenum Publishing Co. Translated from Byulleten Eksperimental'Noi Biologii i Meditsiny, **89**, 3, 318-320.

Taylor, M.J., Duggleby, C.J. and Atkinson, T. (1986) J. Theor. Biol., **121**, 293-320.

Terman, D.S., Young, J.B., Shearer, W.T., Ayus, C., Lehane, Mattioli, C., Espada, R., Howell, J.F., Yamamoto, T., Zaleski, H.I., Miller, L., Frommer, P., Feldman, L., Henry, J.R., Tillquist, R., Cook, G. and Daskal, Y. (1981). N. Eng. J. Med., **305**, 1195-1200.

Tommassen, J. and Lutenberg, B. (1982). Ann. Microbiol. (Inst. Pasteur), **133A**, 243-249.

Towbin, H., Stachelin, T. and Gordon, J. (1979). Proc. Natl. Acad. Sci., USA, **76**, 4350-4354.

Travers, A.A. (1986). Biochem. Soc. Trans., **14**, 199-200.

Tyler-Cross, R., Roberts, C.H. and Chlebowski, J.F. (1989). J. Biol. Chem., **264**, 4523-4528.

Uhlen, M., and Moks, T. (1990). Meths. Enzymol. **185**, 129-143.

Uhlen, M., Nilsson, B., Guss, B., Lindberg, M., Gatenbeck, S. and Philipson, L. (1983). Gene, **23**, 369-378.

Uhlen, M., Guss, B., Nilsson, B., Gatenbeck, S., Philipson, L and Lindberg, M. (1984a). J. of Biol. Chem., **259** 1695-1707.

Uhlen, M., Guss, B., Nilsson, B., Gotz, T. and Lingberg, M. (1984b). J. Bact., **159**, 713-719.

Uhlen, M., Lindberg, M. and Philipson, L (1984c). Immunology Today, **5**, 244-248.

Valerie, K., Fronko, G., Long, W., Henderson, E.E., Nilsson, B., Uhlen, M. and de Riel, J.K. (1987). Gene, **58**, 99-107.

Vieira, J. and Messing, J. (1982). Gene, **19**, 259-268.

Verwey, W.F. (1940). J. Exp. Med., **71**, 635-644.

Vold, B. (1985). Microbiol. Rev., **49**, 71-80.

Von Heijne, G (1981). Eur. J. Biochem., **120**. 275-278.

Von Heijne, G (1983). FEBS, **163**, 17-21.

Von Heijne, G. (1984a). J. Mol.Biol., **173**, 243-251.

Von Heijne G. (1984b). EMBO J., **3**, 2315-2318.

Von Heijne, G. (1985). J. Mol. Biol., **184**, 99-105.

Wahl, G.M., Stern, N. and Stark, G.R. (1979). Proc. Natl. Acad. Sci., USA, **76**, 3683-3687.

Wain-Hobson, S., Nussinov, R., Brown, R.J. and Sussman, J.L. (1981). Gene, **13**, 355-364.

Walker, J.M. (1984). Methods in Molecular Biology, Vol. 2, Nucleic Acids. Humana Press, Clifton, New Jersey, USA.

Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T. and Itakura, K. (1979). Nucleic Acids Res., **11**, 3543-3557.

Wang, P-Z. and Novick, R.P. (1987). J. Bact., **169**, 1763-1766.

Wanner, B.L. and Latterell, P. (1980). Genetics, **96**, 353-366.

Warner, N.L. (1974). Adv. Immunol., **19**, 67-216.

Warnes, A. (1989). PhD Thesis. University of Aston, Birmingham.

Watson, M.G.G. (1984). Nucleic Acids Res., **12**, 5145-5164.

Watts, C., Wickner, W. and Zimmermann, R. (1983). Proc. Natl. Acad. Sci., USA, **80**, 2809-2813.

Weickert, M.J. and Chambliss, G.H. (1990). Proc. Natl. Acad. Sci., USA, **87**, 6238-6242.

Weis, J.H., Enquist, L.W., Salstrom, J.J. and Watson, R.J. (1983). Nature, **302**, 72-74.

Williams, D.C., Van Frank, R.M., Muta, W.L. and Burnett, J.P. (1982). Science, **215**, 687-689.

Winblad, S. and Ericson, C. (1973). Acta. Pathol. Microbiol. Scand. Sect. B., **81**, 150-156.

Wolfe, P.B., Wickner, W. and Goodman, J.M. (1983). J. Biol. Chem., **258**, 12073-12080.

Wyckoff, H.W., Handschumacher, M., Murthy, H.M.K. and Sowadski, J.M. (1983). Advances in Enzymology, **55**, 453-480.

Yamamoto, T. and Yokota, T. (1982). J. Bacteriol., **150**, 1482-1484.

Yoshimura, K., Ikenaka, Y., Murai, M., Tanabe, M., Seki, T. and Oshima, Y. (1983). Gene, **2**, 255-263.

Young, F.E. (1980). J. Gen. Micro., **119**, 1-15.

Zenno, S. and Inouye, S. (1990). Biochem. Biophys. Res. Comm., **171**, 169-174.

Zimmerman, S.E., Smith, F.P., Phillips, T.M., Coffey, R.J. and Schein, P.S. (1982). Br. Med. J., **284**, 1432-1434.

Zoller, M., Atkinson, T. and Smith, M. (1983). Manual For Advanced Techniques in Molecular Cloning Course. Cold Spring Harbour.

Zikan, J., Sima, P., Prokesova, L. and Hadge, D. (1980). Folia. Biol., (Prague), **26**, 261-266.

APPENDIX

Appendix I

Relationship Between Optical Density and Cell dry Weight

In order to accurately assess the level of expression of SpA from different *E.coli* [SpA] plasmid cultures it is necessary to know exactly how much SpA is produced per volume of cells. Determination of the SpA levels produced for a certain dry weight will enable the comparison of SpA levels between the cultures those optical density were measured at 450 nm (localisation experiments) with those measured at 600 nm (later 500 ml culture runs) and will also allow SpA to be expressed as $\mu g\ SpA.mg^{-1}$ cells. To construct a graph of dry weight against optical density a calibration curve of dry weight against % cell suspension must first be constructed.

method

When measuring dry weights it is important to pre-dry the weighing bottles to prevent negative results. Bijou bottles were placed in a 150°C oven to dry. They were weighted beforehand and were left to dry until a constant weight was attained.

Stationary phase JM83 [pPA31] cells (100 ml volume) were harvested (centrifuged 6000 r.p.m for 10 minutes using a SS24 Sorvall rotor) and resuspended in 35 mls of 2xYT. This was deemed a 100% suspensions, 5 mls of the following suspension were prepared by dilution of this: 80%, 60%, 40% and 20%. 1 ml of each suspension was diluted in 9 ml of 2xYT and the optical density at 600 nm and 450 nm measured. 2 ml of each suspension was added to the pre-weighted bottles (4 duplicate samples) and allowed to dry at 115°C until a constant weight was achieved (four days), 2 mls of 2xYT was also added to a pre-weighed bottle and dried down to determine the weight of the medium used.

A sample of stationary phase cells from JM83 [pPA163] and JM83 [pPA34] grown under identical conditions to JM83 [pPA31] was also diluted to the same OD_{600} and OD_{450} as the 100% sample from JM83 [pPA31] and the dry weight determined. No significant difference was seen between the dry weights determined for each plasmid type.

The average dry weight of the cell culture was plotted against the percentage suspension (Figure A1.i). The true dry weight can be calculated from the plot by measuring the distance from the intercept of the slope with the Y axis (mg dry weight).

The resulting values of cell dry weight were calculated for 1 ml of cells and divided by 10 to allow for the 1 in 10 dilution carried in obtaining the optical density of the culture.

10xdil % Suspension	Dry Weight mg ml^{-1}	OD_{600}	OD_{450}
100	0.420	1.238	1.176
20	0.335	1.048	0.972
60	0.250	0.872	0.793
40	0.170	0.584	0.516
20	0.085	0.313	0.282

A graph of optical density versus dry weight for JM83 [pPA31] was plotted (Figure A1.ii). This showed a linear relationship to exist between dry weight and optical density for absorbance wavelength 450 nm and 600 nm until an optical density of 1.0

All small scale culture samples were dilated to give an OD of 1.0 or less and the dilution taken into account in the subsequent calculation of mg SpA mg^{-1} cells.

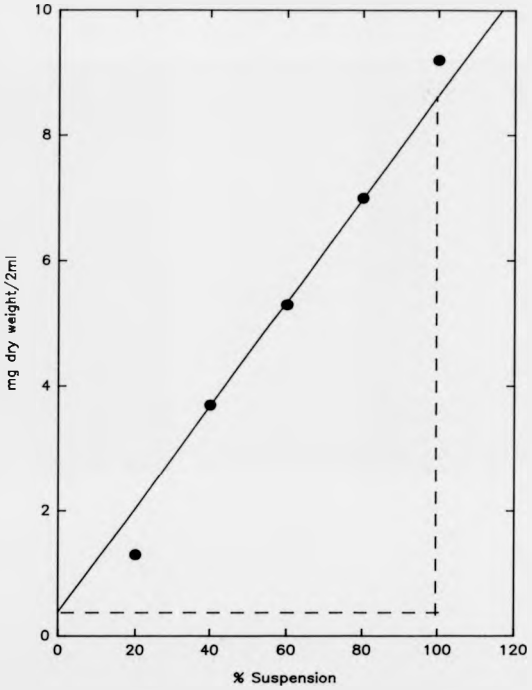


Figure A1.i Calibration curve of cell dry weight vs % suspension to determine actual 100% level

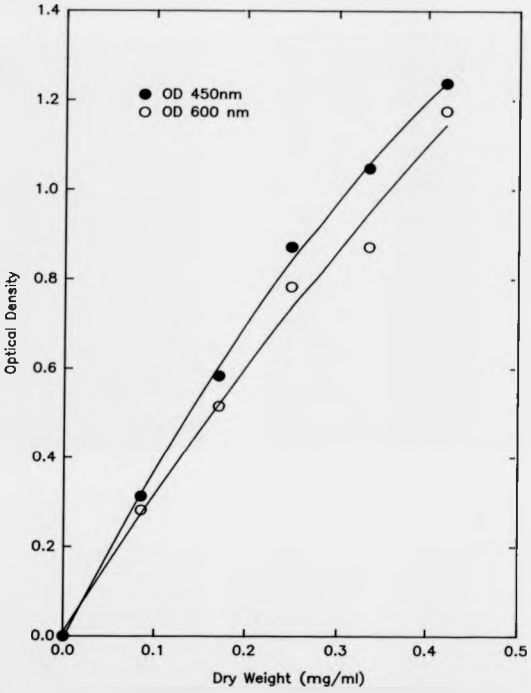


Figure A1.ii Optical density vs dry weight for JM83[pPA31] grown in 2 x YT broth

Appendix II

Materials and Methods of Pilot Plant Fermentation of *E.coli* JM83 [SpA] of Strains

Fermentation of JM83 [SpA] in the 20 litre seed vessel and the 150 litre and 400 litre main culture vessels followed by harvesting of the cells were undertaken by shift technicians at the CAMR Fermentation Pilot Plant under the guidance of Dr. S. Chambers. Standard protocols and quality control procedures, used at the Pilot Plant were followed throughout.

Medium

20, 150 and 400 litre culture medium was made up as follows:

	Per Litre
K_2HPO_4	4.0 g
KH_2PO_4	1.0 g
NH_4Cl	1.0 g
$CaCl_2.2HO$	0.01 g
K_2SO_4	2.6 g
*Trace elements	10 mls
Cusamino acids	20.0 g
Yeast extract	3.0 g
Glycerol	46.0 g

*Trace elements: 5 g of EDTA was dissolved in 300 ml of demineralised H_2O then each of the following (dissolved separately in a small volume of demineralised H_2O) was added in the order shown.

$FeCl_3$	0.5 g
ZnO	0.05 g
$CuCl_2.2H_2O$	0.01 g
$CoNO_3.6H_2O$	0.01 g
$(NH_4)_6\ Mo_7O_{24}.4H_2)$	0.01 g

these were made up to a final volume of 1 litre and added to give a 1% final concentration. The pH of the medium was brought to pH 7.0 using NaOH or H_3PO4. The medium was sterilised by heating up to 121^{o}C for 30 minutes at 20 PSI.

1 M Magnesium chloride -1 ml/litre and ampicillin to give final 50 $\mu g\ ml^{-1}$ were added via a sterile filter first prior to inoculation.

Culture Conditions

The culture vessel and associated equipment used were of the design described by Melling and Scott (1972). The pH was monitored and controlled between present parameters by automatic additions of either 20% H_3PO_4 or 3.2N NaOH. Foaming was controlled by the addition of Mazu/8005 (3 mls in 20 litre cultures, 55 mls in 150 litre culture and 15 mls in 400 litre culture) to the growth medium prior to sterilisation. Temperature was monitored at 37^{o}C by a thermostatically controlled water jacket. The Co_2 content of effluent gas from the culture was monitored using a Grubb Parsons model IRG-A-20 CO_2 analyser. The culture was aerated with presterilised bottom air at a rate of 12 L min^{-1} per 20 litre culture, 75 L min^{-1} per 150 litre and 300 L min^{-1} per 400 litre culture. The culture was stirred at 250 rpm. Samples were taken from the culture at hourly intervals for routine monitoring, purposes (100ml discard). The pH and OD_{600} using a Pye Unicam SpB-100 spectrophotometer were monitored every hour. Cell growth was also monitored every 2 hours by lab dry weight (LDW) in conjunction with viable counts (made on nutrient agar plates (either L-agar or 2xYT agar). Plasmid loss could also be monitored by plating onto Ampicllin (50 $\mu g\ ml^{-1}$ containing nutrient agar plates and comparing the viable count to that of the non-ampicillin plates. Purity of culture was monitored every 2 hours by random selection of a few colonies from the viable count plate, these were then scanned under the light microscope and tested for known biochemical characteristics of the parent strain. The SpA levels within the cells and the culture supernatant were determined at 4 hours per

inoculation and then hourly, using the monorocket SpA assay described in section (2.2.14.2). All the above readings were also carried out on a final sample taken immediately prior to cell transfer to a larger fermenter or prior to harvesting.

Appendix III

S. aureus Protein A Gene Fusions

The following protein A gene fusions have been constructed to date. The first protein of each fusion indicates the 5′ end of the gene fusion and the second protein the 3′ end of the gene fusion.

SpA - Protein A - β-galactosidase (Uhlen, *et al.*, 1983; Uhlen *et al.*, 1984c; Nilsson *et al.*, 1985a; Hellebust *et al.*, 1987; Hellebust, *et al.*, 1988).

SpA - Insulin like growth Factor I (Nilsson *et al.*, 1985b)

SpA - antigenic peptide, SpA acts as an adjuvant (Lowenadler *et al.*, 1986; Lowenadler *et al.*, 1987, Valerie *et al.*, 1987).

β-glucuronidase - SpA (Anilonis and Palmer, 1987).

SpA - Ricin A (Kim and Weaver, 1988)

SpA - Carboxypeptidase G2 (Shuttleworth, unpublished results)

SpA - Alkaline phosphatase (Nilsson and Abrahmsen, 1990)

SpA - Bovine pancreatic trypsin inhibitor (Nilsson and Abrahmsen, 1990)

Metapyrocatechase - SpA (Kobatake, *et al.*, 1990)

SpA - Aequorin (Zenno and Inouye, 1990)

SpA - Glutathione S-Transferase (Lew *et al.*, 1991)

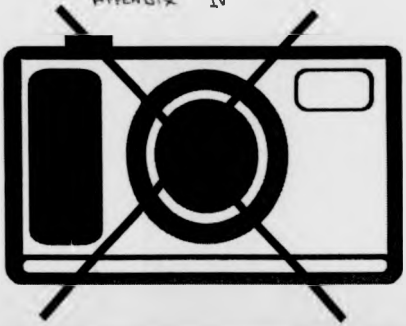
SpA - Luciferase (Lindbladh *et al.*, 1991)

SpA - β-lactamase (Georgiou and Baneyx 1990)

Appendix IV Publications

PUBLISHED PAPERS NOT FILMED FOR COPYRIGHT REASONS

APPENDIX IV



NUMEROUS ORIGINALS
IN COLOUR

VARIABLE PRINT QUALITY